Risk-benefit and Quality-of-Life Analyses of Prescription Drugs

Health and Welfare Canada Santé et Bien-être social Canada

Canada

RISK-BENEFIT AND QUALITY-OF-LIFE ANALYSES OF PRESCRIPTION DRUGS

George Torrance, Ph.D.[1,2,3]
Barbara Markham, M.B.A.[1,2]
David Feeny, Ph.D.[1,2,4]
David Rosenbloom, Pharm.D.[5,6]
Amiram Gafni, Ph.D.[1,2]

1 Centre for Health Economics and Policy Analysis, McMaster University
2 Department of Clinical Epidemiology and Biostatistics, McMaster University
3 Faculty of Business, McMaster University
4 Department of Economics, McMaster University
5 Department of Medicine, McMaster University
6 Department of Pharmaceutical Services, Chedoke-McMaster Hospitals

Health and Welfare Canada
May 1990

90-05875/

© Minister of Supply and Services Canada
1990
Catalogue No. H42-2/24-1990
ISBN 0-660-55555-7

ABSTRACT

A review of the literature on approaches to risk-benefit and quality-of-life analyses of prescription drugs was conducted. The use of quality-of-life measurement techniques in the clinical trials for certain new drugs are encouraged. Utility measurement in particular is endorsed as a powerful technique for determining the balance of risks versus benefits for many new medicines.

TABLE OF CONTENTS

Page

Abstract iii

List of Figures vii

List of Tables vii

1. EXECUTIVE SUMMARY 1

2. INTRODUCTION 4

3. RISK-BENEFIT ANALYSIS FOR NEW DRUGS 5

3.1 The Drug Assessment Iterative Loop 5
3.1.1 Scope of This Report 8

3.2 Risk, Benefit, and Risk-Benefit Analyses: Definitions 9
3.2.1 Risk (Safety) 9
3.2.2 Benefit (Efficacy/Effectiveness/Efficiency) 10
3.2.3 Risk-Benefit Analysis 10

3.3 Methods of Measuring Risk-Benefit 11
3.3.1 Identification 11
3.3.2 Quantification 12
3.3.3 Evaluation 12

3.4 New Prescription Drug Licensing 14
3.4.1 Review of Canadian Licensing Requirements 14
3.4.2 Current Approaches by Drug Companies 15

4. LIMITATIONS OF THE CURRENT SYSTEM FOR ASSESSING RISK BENEFIT OF NEW DRUGS 17

4.1 RCCT Sample Size Limitations 17

4.2 RCCTs Usually Do Not Predict Long-Term Effects 17

4.3 Statistical vs. Clinical Significance 18

4.4 What is "Clinical Significance"? 18

4.5 Difficulty of Assessing Multiple Effects 18

... continued

TABLE OF CONTENTS (cont'd)

Page

5. QUALITY OF LIFE IN THE ASSESSMENT OF PRESCRIPTION DRUGS 20

5.1 Definition of Health-Related Quality of Life 20

5.2 Approaches to Measuring QOL 21

5.3 Definition of "Utility" 22

5.4 Utility Measurement 23

6. EXPLORING THE USE OF UTILITY-BASED QOL MEASUREMENT IN THE ASSESSMENT OF NEW DRUGS FOR LICENSING PURPOSES 26

6.1 Advantages of Employing Utility Measures 26

6.1.1 Generalizability 26
6.1.2 Comprehensiveness 26
6.1.3 Integration of Morbidity and Mortality Effects 26
6.1.4 Capturing of Two Dimensions of Risk 27
6.1.5 Integration of Risks and Benefits 27
6.1.6 Preference Measurement for Hypothetical and Actual Health States 27
6.1.7 Representation of Multiple Viewpoints 27
6.1.8 Incorporation of Time and Risk Preferences 28
6.1.9 Rigorous Theoretical Foundation 28
6.1.10 Integration with Economic Evaluation Measures 28
6.1.11 Integration with Clinical Trials 28
6.1.12 Usefulness for Prescribing Physicians 28

6.2 Disadvantages of Utility Approach to Measuring Quality of Life . 29

6.3 An Example of the Incorporation of Quality-of-Life Measurement in a Drug Trial 29

7. SUMMARY AND CONCLUSIONS 33

8. FINALE 35

9. REFERENCES 36

Appendix A: Selected Major Recent Conferences on the Measurement of Health-related Quality of Life 41

Appendix B: Major Bibliographic Resources 42

LIST OF FIGURES

		Page
Figure 1:	The benefit-risk matrix	43
Figure 2:	The drug assessment iterative loop	44
Figure 3:	Feeling thermometer approach for ordinal ranking	45
Figure 4:	Standard gamble approach for eliciting utility value for a chronic state	46
Figure 5:	Time trade off approach for eliciting utility value for a chronic state	47

LIST OF TABLES

		Page
Table 1:	Selected Definitions of "Efficacy"	48
Table 2:	Taxonomy of Quality-of-Life Assessment Measures	49
Table 3:	Example of Health-State Description	50
Table 4:	The Advantages and Disadvantages of Utility Measures	51
Table 5:	Changes in Clinical and Health Status Measures and Composite Scores between Baseline and Sixth Month By Treatment Group (n = 303)	52
Table 6:	Results of Utility and Other Measures	53

1. EXECUTIVE SUMMARY

Virtually all beneficial drugs impose some degree of risk. How can one determine which drugs should be made available to the public? These decisions require an assessment of the risks and benefits to human health, a process often called "risk-benefit analysis". This report discusses current and potential approaches to the measurement of the risks and benefits of new prescription drugs.

Risk combines the probability of an adverse event with the severity or harm to human health if the event occurs. *Safety* is the converse of risk; a drug is safe if its risk has been judged acceptably low for a specified situation.

Benefit refers to the improvement in human health attributable to the drug. Traditionally, health benefits have been measured as improvements in morbidity and mortality. Morbidity is usually measured by traditional diagnostic criteria. Occasionally some consideration is given to levels of physical, emotional, and social functioning. Mortality is usually measured by death rates, with occasionally some consideration given to longevity.

Risk-benefit analysis is a technique for combining the risks and benefits to human health, comprehensively measured, and reporting these impacts in a way that facilitates judgment about the balance of risks and benefits to the population. The risk-benefit matrix (see Figure 1, page 42) is a useful visualization of the process. Here, risk is represented along one axis, while benefit is represented along the other. Low risk (safe) drugs will be acceptable even at low levels of benefit. On the other hand, high risk drugs may be unacceptable at virtually any level of benefit. The difficulty lies in judging the many drugs that fall into the middle of the matrix.

Health-related quality of life represents a new and comprehensive approach to measuring the health status of individuals and groups. The concept may be used to measure both the benefits of drugs and their risks. In this way, quality-of-life measures have the potential for use as an integrating metric that can bring together the positive impacts (benefits) and the negative impacts (risks) of a new drug to determine if the overall balance is favourable. Moreover, these measures have the potential to determine this both within and across patient populations. This is a powerful extension of risk-benefit analysis.

Overall quality of life is influenced by a variety of factors (income, job satisfaction, social opportunities) in addition to health status. In the context of clinical studies, however, the comparison of alternatives should appropriately focus on that part of the quality of life that is health-related.

A variety of approaches to quality-of-life assessment have emerged. Such a diversity is both inevitable and appropriate. At present there is no gold standard for measurement of quality of life; nor is one likely to develop.

Quality-of-life measures may be classified into two broad groups: specific instruments (e.g., Chronic Respiratory Questionnaire), and generic instruments. Generic instruments are subdivided into health profiles (e.g., Sickness Impact Profile), and utility measures (e.g., Standard Gamble Instrument).

The utility approach is particularly powerful and useful in the assessment of new prescription drugs. *Utility* is a concept used in economics and in decision analysis, which refers to the level of satisfaction or enjoyment experienced by consumers. In health applications, utility refers to the satisfaction that individuals derive from different levels of health status or health outcomes. It also represents the strength of the preferences of individuals for various outcomes.

General methods to obtain cardinal measures of utility for health outcomes are well developed. When carefully customized and deployed the instruments have proven reliable and valid. The two major approaches are the "standard gamble" and "time trade-off". Both are frequently supplemented with the prior use of a visual analog rating scale (feeling thermometer).

The advantages of the utility approach to quality-of-life measurement include its generalizability, comprehensiveness, ability to integrate mortality and morbidity effects, ability to integrate risks and benefits, ability to elicit preferences both for hypothetical and actual situations, ability to represent multiple viewpoints, ability to use different types of evaluators for health states, its incorporation of time and risk preferences of subjects in their health state evaluations, its solid theoretical foundation, its consistency with economic evaluation, and its feasibility for use in clinical trials. The utility approach produces a single score on a generalized scale from 0.0 (dead) to 1.0 (healthy). Information on trade offs among the benefits of treatment and side effects are reflected in the utility scores. Utility analyses can also be readily combined with pecuniary (dollar) measures of costs and benefits of health care alternatives to produce cost-utility analyses.

The advantages of the utility approach do not come without disadvantages. Utility measures are not highly precise at the individual level, and require repeated measures or group applications to achieve high levels of precision. Utility measures require labour intensive interviews conducted by trained professional interviewers. Utility scores are not always easy to interpret. These instruments cannot be applied

to all subjects. Finally, utility scores do not reveal why a particular health outcome is highly or lowly valued; supplementary measures are required to tap the detailed reasons.

Utility and quality-of-life measures have been used successfully in clinical trials. For example, a broad array of quality-of-life measures were incorporated in a recent international multi-centred trial of a new pharmacological agent, Auranofin, for the treatment of arthritis. A utility measure was among the quality-of-life instruments used, and it was especially responsive. The Auranofin example is presented in some detail in this report to illustrate the potential of the approach.

In sum, methodologies for measuring health-related quality of life, including the utility approach, are now sufficiently well developed that it is scientifically appropriate for them to be used in drug trials. Quality-of-life measures add important and useful information to that which is traditionally submitted for drug licensing. The utility approach to quality-of-life measurement is particularly powerful, appealing, and deserving of consideration.

2. INTRODUCTION

Given that virtually all beneficial drugs also impose some degree of risk, how can one determine which drugs should be made available to the public? These decisions require an assessment of the risks and benefits, a process often called "risk-benefit analysis" (RBA). This report discusses current and potential approaches to the measurement of risks and benefits of new prescription drugs. The point of view is that of Canadian society. The Health Protection Branch (HPB) of Health and Welfare Canada represents society's interests in assuring that drugs are safe and efficacious.

Quality of life is an important health outcome. Over the last few years there has been a dramatic increase in activity worldwide in developing and applying methods that measure quality of life (for example, see Appendix A for a list of recent conferences devoted to the topic and Appendix B for details on a U.S. Government Clearinghouse that has a major focus in this area). As a result of this activity, a number of quality-of-life measurement techniques, including the utility approach, are now relatively well developed.

New prescription drugs can impact on quality of life in positive ways through a beneficial treatment effect, and negative ways through side effects and adverse drug reactions. Utility-based quality-of-life measures have the potential to measure the positive and negative impacts in the same units, and thus facilitate a balancing of risks against benefits. This report explores this possibility.

The material in the report is organized as follows. First, an idealized algorithm, the Drug Assessment Iterative Loop, is used to describe the issues surrounding the evaluation of drugs, as well as the mandate of the HPB and the scope of the report. Then, safety and efficacy are discussed in terms of risk, benefit, and risk-benefit analysis (RBA). Techniques for risk assessment are discussed. Recently developed quality-of-life measurement techniques are then reviewed. Finally, the report investigates the application of the utility approach to the measurement of health-related quality of life as a possible method to assist in the balancing of risks and benefits. The approach is illustrated through an example from the clinical trial of an oral gold compound (Auranofin) for arthritis patients.

3. RISK-BENEFIT ANALYSIS FOR NEW DRUGS

3.1 The Drug Assessment Iterative Loop

The point of view for this report will be that of Canadian society. In Canada, the Health Protection Branch (HPB) of the Department of Health and Welfare[1] has the responsibility[2] for assuring the safety and efficacy of new drugs. We can position the role of the HPB in assessing the safety, efficacy, and effectiveness of new drugs from a societal viewpoint, by employing an adapted version of the Measurement Iterative Loop (Tugwell et al. 1985) and the Technology Assessment Iterative Loop (Tugwell et al. 1986) as shown in Figure 2 (page 43). This loop is a simple, idealized model useful for assembling and organizing information. In addition to describing the mandate of the HPB, the Drug Assessment Iterative Loop will be used to delineate the scope of the report.

The Drug Assessment Iterative Loop provides a systematic approach to the assembly of specific subsets of health information necessary for rational decision making on the adoption and utilization of new drugs. This framework subdivides the spectrum of information into subgroups that constitute a logical progression from quantifying the burden of illness to validating that the drug is safe and prevents or ameliorates the illness, to diffusion of the drug, and coming full circle, to determining whether the burden of illness has been reduced. Each of the steps in the Drug Assessment Loop poses different research or evaluation questions. For example, the framework distinguishes the concepts of drug safety, efficacy, effectiveness, and efficiency. The loop format emphasizes the importance of monitoring a drug after it has been approved and marketed. The process of reducing the burden of illness is also iterative. In most health care situations, the burden of illness is reduced only by increments and repeated adjustments are needed to relieve even that portion for which beneficial interventions exist (Tugwell et al. 1985, 1986).

---

[1] This report specifically addresses the mandate of the Bureau of Human Prescription Drugs of the Drugs Directorate, although much of the discussion can be applied to other bureaux and directorates as well.

[2] *Narcotic Control Act* and Regulations and *Canadian Food and Drugs Act* and Regulations.

The mandate of the HPB is focused on step 3 (safety and efficacy) of the loop but also encompasses step 2 (screening and diagnostic accuracy), and step 4 (community effectiveness). The HPB is also concerned to some extent with step 7 (monitoring and reassessment). It is not the mandate of the HPB to consider drug efficiency (i.e., costs of a drug compared to alternatives), although this may be of concern to particular (i.e., hospital or provincial) drug formularies, or individual patients or physicians once the drug has been approved for marketing by the HPB.

The seven steps of the Drug Assessment Iterative Loop are:

- **Step 1: Burden of Illness**

  When appraising a new drug from a societal viewpoint the first step is to determine the burden of illness that the drug is targeted to relieve. The burden of illness may be assessed by determining current levels of morbidity (e.g., physical function, emotional function, social function, symptoms, and disability days) or mortality (e.g., death rates and years of life lost). The measurement tools used to determine the burden of illness should be relevant to the illness, accurate, and easy to use and interpret. In addition to assessing the burden of illness, it is desirable to understand what causes the illness and then to determine if the illness is modifiable by drug therapy. Evidence for the causes of health problems contributing to the burden of illness can be assessed by following guidelines such as those developed at McMaster University by the Department of Clinical Epidemiology and Biostatistics (1981).

- **Step 2: Screening and Diagnostic Accuracy**

  The second step focuses on establishing evidence of specificity, sensitivity, and utilization of specific tests and drugs to diagnose the illness. This includes the issue of specificity and sensitivity of screening programs and other exercises conducted by healthcare professionals in determining who might benefit from drug therapy and who might not.

- **Step 3: Safety and Efficacy**

  This step is concerned with whether the drug can work. Does it do more good than harm for patients, under ideal conditions (i.e., optimal diagnostic accuracy and full health care provider and patient compliance with associated recommendations)?

- **Step 4: Community Effectiveness**

  Effectiveness is concerned with usefulness. Does the drug do more good than harm to people to whom it is offered under real life conditions? This consideration not only encompasses diagnostic accuracy (step 2 above) and drug efficacy (step 3 above) but also health care provider compliance, patient compliance, and coverage. Health care provider compliance is concerned with the appropriate diagnostic and management actions of the health care provider (compliance with recommendations for use of the drug). Patient compliance is concerned with whether or not the patient complies with the health care providers recommendations and treatment. Ease of taking a drug, side effects attributable to the drug, cost of the drug to the patient, or extent to which the patient may understand the potential benefits of taking the drug — all affect patient compliance.

  Coverage refers to the extent to which the drug is being appropriately utilized by all those who could benefit from it. Utilization depends on availability (easy to obtain?) and acceptability (of resources, facilities, health professionals behaviour, and expected benefits of health service). In order to secure acceptable coverage, the drug therapy program must be related to the need for it in a defined population during a specified period (Tugwell, 1985).

- **Step 5: Efficiency**

  This step is concerned with determining whether an effective drug therapy can be delivered in a manner in which the benefits of the program (e.g., number of lives or life-years saved, disability days averted, improvement in quality of life, reduction in future treatment costs) outweigh the costs (e.g., costs of screening, treatment and side effects; impact of side effects on quality of life).

  Cost-benefit, cost-effectiveness, and cost-utility methods for measuring efficiency are discussed in Drummond et al. (1987).

  Unlike safety, efficacy, and effectiveness, the assessment of drug therapy efficiency is clearly not within the mandate of the HPB. To measure efficiency, we compare the incremental costs and incremental benefits of the drug to an alternative treatment for the same problem and to alternative uses of resources. The efficiency of a drug therapy is not static. Drug therapy must be reassessed relative to new drugs, technologies, and changing environments. Efficiency is of concern generally to society and

specifically to those who might have to pay for the new drug therapy (i.e., individual patients, hospitals, provincial, and private drug formularies). Efficiency considerations are beyond the scope of this report.

- **Step 6: Synthesis and Implementation**

  The sixth step integrates feasibility with the estimates of community effectiveness and efficiency obtained in the previous steps to make recommendations for implementing an intervention. This step identifies possible limiting constraints on the effectiveness and efficiency estimates currently existing in the setting where the intervention will be implemented. It identifies whether and to what extent these constraints can be removed or reduced, and comes to a conclusion about the likely impact of the drug on the burden of illness. Constraints include social, cultural, and political barriers as well as adequacy and availability of facilities and manpower and budgetary considerations. Inherent in this step is the setting of realistic objectives and evaluation criteria (Tugwell, 1985).

- **Step 7: Monitoring and Reassessment**

  This step assesses the short-term, intermediate, and long-term success of the new drug therapy in reducing the burden of illness. The selection of markers to monitor the success of the new drug needs consider the spectrum of illness and age groups, accuracy and feasibility (Tugwell, 1985).

  The Drug Assessment Iterative Loop is completed by returning to the hopefully reduced burden of illness. Each circuit of the loop provides information on how the drug can reduce further the burden of illness.

3.1.1 Scope of This Report

The HPB is concerned directly with a number of steps in the loop, namely step 2 (diagnostic accuracy), step 3 (safety and efficacy), step 4 (community effectiveness), and step 7 (monitoring and reassessment). The HPB does assess the safety of substances used in the diagnosis of illness in patients, such as radio-opaque dyes for radiology. An important part of the evidence submitted to the HPB by pharmaceutical firms comes from Phase III trials conducted in realistic settings, thus providing evidence for step 4 (community effectiveness). Post-marketing surveillance, such as the Federal Voluntary Drug Adverse Reaction Reporting Program (DARRP), falls into step 7. The licensing decision, however, rests largely on the efficacy trials, the evidence for step 3, and the primary focus of HPB (safety and efficacy). The remainder of this report will focus on the role of the HPB at steps 3 and 4 in assessing the risk benefit of new drug therapies.

3.2 Risk, Benefit, and Risk-Benefit Analyses: Definitions

3.2.1 Risk (Safety)

Public agencies involved with risk regulation tend to prefer the positive term *safety* to the negative term *risk*. However, safety and risk are two sides of the same coin, and both are relative concepts (i.e., no drug is ever completely safe or risk-free). Intuitively, a safe drug causes no harm. Here the term 'safe' is used to refer to a level of risk which has been judged to be so small that it is insignificant (O'Brien, 1986).

As defined in *Webster's New World Dictionary* (1980), risk is:

> ".... the chance of injury, damage, or loss; dangerous chance; hazard;"

Risk can be thought of as a concept that combines the "probability" or chance of the occurrence of an event and the "severity" or harm to human health if the event should occur (Lowrance, 1976, p. 94). See Appendix B for reference to a recent comprehensive bibliography on risk. The concept of risk implies that investigators and policy makers should be concerned with both the nature of the risk and the probability of its occurrence. The Royal Society of Great Britain (1983) definition of risk is consistent with this view.

According to the Society, risk is:

> ".... the probability that a particular adverse event occurs during a stated period of time, or results from a particular challenge."

The Society's definition includes the adjective "adverse", emphasizing that the event is undesirable, having connotations of potential loss, harm, injury or damage.

A more explicit definition of risk comes from the Office of Technology Assessment (1978):

> ".... the probability of adverse or untoward outcomes occurring and the severity of the resultant harm to health of individuals in a defined population, associated with the use of a medical technology applied for a given medical problem under specified conditions of use."

This definition can be applied to drug evaluation as well. It is comprehensive in that it includes several factors: the medical problem for which the technology (or drug) being evaluated is applied, the population affected, and the conditions of use under which the technology (or drug) is applied. The definition covers risk under

ideal (research) settings, and under average or typical settings (community). Normally, when "efficacy and safety" judgments are being discussed, risk is assumed to be measured under ideal conditions of use.

Given the above definition of risk, safety can be specified as:

> ".... a judgement of the acceptability of risk in a specified situation (Banta 1981, p. 99)."

3.2.2 Benefit (Efficacy/Effectiveness/Efficiency)

The HPB requires more than the demonstration of safety. A drug must also be shown to be efficacious, either directly or indirectly, before the HPB will license it for marketing. Like safety, efficacy is a relative concept for which there is no shortage of definitions (Table 1, page 47). According to Banta (1981), there are four factors, which when taken together, form a comprehensive view of the concept.

The factors are:

a) benefit to be achieved,

b) medical problem giving rise to the use of the technology (drug, etc.),

c) population affected, and

d) conditions of use under which the technology is applied.

The term "benefit" is related to the terms "efficacy" (benefit under ideal conditions of use), "effectiveness" (benefit under average conditions of use), and efficiency (benefit compared to cost).

The outcome criteria used to measure benefit have usually been restricted to measurement of morbidity and mortality; less consideration has been given to life expectancy (longevity) or psychosocial and functional factors. The definition of benefit will depend on the goals of the investigator and the type of drug being assessed. For a recent discussion on measures of clinical benefit, see Laupacis, Sackett and Roberts (1988). They review the advantages and shortcomings of such yardsticks as relative risk reduction, odds ratio, absolute risk reduction and number needed to be treated from the perspective of a clinician assessing individual risk.

3.2.3 Risk-Benefit Analysis

In 1971, Feinstein indicated there is no satisfactory standard procedures for assessing the safety and efficacy of drugs. Today, the terminology of risk-benefit analysis, more generally known as "risk evaluation", is still far from standardized (Jones and Akehurst, 1980).

The Treasury Board of Canada (1979, p. 10) defines RBA as:

> ".... a special case of cost-benefit analysis (CBA) where explicit attention is devoted to categories of human costs (risk of loss of life or limb) for which established techniques of evaluation do not yet exist. In risk-benefit analysis, benefits should be interpreted in a net sense (i.e., the balance of the ordinary benefits over the ordinary costs), but the risk of loss of life and of human suffering represent extraordinary costs, to be evaluated separately."

In this definition the term "costs" refers not to pecuniary or dollar costs but to "human" costs (i.e., to increases in probability of reductions in health status). Thus, risk-benefit analysis is a method to analyze comprehensively all impacts on human health, both positive and negative, and to report these impacts in a way that facilitates reaching a judgment about the balance of risks and benefits to the population.

Teeling-Smith's benefit-risk matrix, shown in Figure 1 (page 42), provides a useful visual representation of the problem. It shows a matrix, with safety along the vertical axis and clinical and social significance along the horizontal axis. A medicine lying in the bottom part of the matrix is acceptable because it is safe while a medicine at the upper left corner would be unacceptable because it provides little benefit at great risk. The difficulty lies in determining how to position medicines in this matrix. Complicating the picture are drugs that can shift around the matrix depending on how they are used. Chloramphenicol, for example, falls in the top right-hand corner when used against typhoid but shifts to the top left-hand corner if prescribed for a sore throat. A medicine such as Benoxaprofen that is safe and important for young crippled arthritics may not be safe if given to an elderly arthritic patient (Teeling-Smith, 1983).

3.3 Methods of Measuring Risk Benefit

Jones and Akehurst (1980) discuss the procedure of risk assessment as consisting of several stages: risk identification, quantification, and evaluation. The following material discusses these stages in risk-benefit analysis.

3.3.1 Identification

"Identification" (the first stage) of the set of risk events associated with some set of circumstances, is relatively straightforward. A list of all possible risk events is all that is required. Nonetheless, the details of the list compiled depends on the compiler's perception of risks. Furthermore, the risks of some new drugs may not have been identified yet.

3.3.2 Quantification

The next stage, "risk quantification", presents two difficulties. First, the risk events may be complex and the estimation of several interrelated risk probabilities may be necessary. Multiple effects may result in disparities or contraindications. Secondly, there may be few relevant data on which to build the estimation of probabilities.

The usual approach to "risk quantification" is to tabulate and compare the numbers of side effects or adverse reactions that occur in patients treated with test drugs or placebo. Laupacis, Sackett and Roberts (1988) point out that, in the absence of quality-of-life assessments that reveal both the benefits and harms of treatments in all patients, absolute harm from therapy can be quantified in the same way as absolute benefit — by subtracting the harm in the control group from that in the treatment group. However, these methods do not address the issue of how much of an adverse effect is cause for concern? How much of a beneficial effect is needed to establish efficacy? In the case of adverse effects, conservative "worst case" assumptions are used, but how conservative is conservative? Here the assumptions are not explicit. Operational criteria to identify adverse reactions are not often provided (Feinstein, 1971).

Another approach to the problem of multiple effects is a decision tree analysis. The "tree" or diagram breaks down complex hazards into a series of simpler contingencies for which estimates of probability of occurrence are available in principle. Fineberg (1980) and Weinstein and Fineberg (1980) describe this approach.

The second problem — a lack of suitable data on which to base estimates of the risk probabilities — is most severe when rare events are being considered. In most risk quantification exercises, human subjective integration, usually judgments of "experts", will be incorporated to a substantial extent (Jones and Akehurst, 1980). Formal techniques such as "probability encoding" (Spetzler and von Holstein, 1975) exist for quantifying the probability judgments of experts in such circumstances. Bayesian methods may also be used in risk estimation exercises (Altshuler, 1976). Finally, techniques such as the Delphi technique (Linstone and Twoff, 1975) can be used for reconciling differences in these judgments.

3.3.3 Evaluation

The third stage of risk assessment is risk evaluation or risk-benefit analysis. Individuals (explicitly or implicitly) weigh estimates of risks against benefits when making decisions about their future behaviour. Such assessments are also made by society, through agents of the public trust such as the HPB.

Although there is extensive literature on RBA, there is a lack of examples where analytic methods have been used as a basis for licensing decisions. A review of the literature indicates a need to separate scientific versus value judgments.

Therapeutic indices have been used to describe, scientifically, the relationship between the doses of a drug required to produce undesired and desired effects. The therapeutic index is usually defined as the ratio between the median toxic dose and the median effective dose (TD50/ED50), or the median lethal dose and the median effective dose (LD50/ED50). Goodman and Gilman (1985) noted that a drug does not have a single but many therapeutic indices. The margin of safety of aspirin for relief of headache is greater than the margin of safety for relief of arthritic pain. Similarly, several therapeutic indices can be calculated for each desired effect. A synthetic opiate may cause less constipation than morphine and yet afford no advantage over the parent compound with regard to the respiratory depression or sedation. Moreover, a drug may be selective within one context yet still be nonselective within another. The antihistamines are selective antagonists of histamine, yet only two of these drugs produce this selective peripheral effect without causing significant central sedation. Finally, a drug may have an adequate margin of safety in most patients, but this description is meaningless for the patient who exhibits an unusual response to the drug. For example, penicillin is essentially nontoxic in the great majority of patients, yet it may cause death in those who have become allergic to it. Thus the therapeutic index provides very limited information for licensing purposes.

Banta (1981) describes five techniques for evaluating safety and efficacy:

- **Preclinical** techniques are animal or biochemical tests that provide a guide as to a drug's potential therapeutic activity and toxicity. The accuracy of animal models in determining the probable effects of drugs on humans is a controversial issue. Others have addressed this issue in more detail (Laurence, McLean and Weatherall, 1984; Zbinden, 1982).

- **Informal** techniques are described by Banta to include personal experience, dominated by qualitative impressions; and peer experience, which develops through the exchange of personal communication, journal articles, and professional consensus, either formal or informal (Banta, 1981, p. 104.).

- **Epidemiology** is the study of the determinants and the distribution of diseases and injuries in populations. Epidemiological and statistical methods are frequently used in computer modelling and multiple regression techniques to examine efficacy of medical interventions.

- In **Randomized Controlled Clinical Trials** (RCCTs) subjects are randomly assigned to experimental and control groups. The control group receives a placebo or an established standard treatment. The experimental group receives the new treatment. Disadvantages of RCCTs are discussed in Section 4. The advantages include elimination of bias from the assignment of treatment, prevention of bias with respect to variables that exist in the experiment but are not considered in the design, and validity of statistical tests of significance that are used to compare treatments (Banta 1981, p. 107).
- **Formal synthesis** is employed when evidence from the above techniques does not lead to an unequivocal decision. Synthesis involves a compilation of all available information. An appropriate model here is the Drug Assessment Iterative Loop discussed earlier.

3.4 New Prescription Drug Licensing

This section provides a brief review of the types of studies undertaken and the types of data collected in the process of licensing and monitoring new drugs.

3.4.1 Review of Canadian Licensing Requirements

The following are four phases of clinical testing for a new drug:

- **Phase I** — consists of initial testing of a study drug in normal volunteers or relatively healthy patients to determine safety and pharmacologic effects (e.g., clinically important adverse reactions, physical signs, and/or laboratory abnormalities), dose frequency, and duration of exposure to a drug. Small studies in patients may be performed to assess clinical effectiveness. The total number of subjects to whom the drug is administered is typically about 80.
- **Phase II** — studies are initial studies designed to evaluate drug efficacy in selected populations of patients for whom the drug is intended. Relative safety information is also determined in Phase II studies. The studies seek to ascertain, for example, when a drug is expected to cause a specific change in a laboratory parameter. The total number of patients to whom the drug is administered is usually about 200.

- **Phase III** — studies are controlled and uncontrolled clinical trials conducted in patient populations for which the drug is eventually intended. These studies generate information on both safety and efficacy in relatively large numbers of patients under normal conditions of use. These studies provide much of the information that is needed for the package insert and labelling of the drug. About 500 to 3,000 patients are typically administered the drug in this phase.
- **Phase IV** — postmarketing evaluation studies are often conducted to provide additional details required to learn more about the drug's efficacy and/or safety profile. Different formulations, dosages, durations of treatments, and drug interactions may be evaluated. New groups by age, gender, race, or other types of patients may be studied. They can provide information on rare or infrequent adverse reactions as well as evidence of previously unknown beneficial effects. In Canada, postmarketing surveillance studies are not required for purposes of securing or maintaining a license to market a drug. The company marketing the drug usually must, however, report to the HPB on all Canadian adverse reactions. In practice worldwide information on adverse reactions is shared with the HPB.

3.4.2 Current Approaches by Drug Companies

Because much of the important information for risk-benefit analyses is derived from phase III trials, they are the focus of this section. Most of the discussion, however, applies to other phases of drug development.

In selecting safety and efficacy criteria to use in trials, measures of efficacy that are well established for the illness(s) being studied are typically used. Tests must also be evaluated within the resource constraints of the study and must be shown to be effective and efficient for their purpose. Anticipated use(s) of a drug (whether for topical or parenteral usage, for generally healthy or seriously ill patients, life saving or minimal therapeutic effect) as well as previous experience with the drug (toxicological profile) or similar drugs suggest specific safety tests to incorporate into a study. There is no standardized series of laboratory parameters that are evaluated in all clinical studies, nor is there one standard that is applied to all drugs. There are, however, flexible standards and general guidelines for basic haematology, clinical chemistry, urinalysis, and ocular safety parameters.

There are various types of criteria that may be used to measure efficacy (or safety) parameters:

a) presence or absence criteria (s mptom, sign, lesion, or other manifestation of the illness that is either present or not);

b) graded or scaled criteria (such as visual analogue scales);

c) relative change criteria (measurements in specific tests that provide direct or indirect indications of efficacy); or

d) global criteria (overall evaluation of a patient's illness or change in illness severity based on specific or general criteria or a scoring system to rate each of the factors) (Spilker, 1984).

4. LIMITATIONS OF THE CURRENT SYSTEM FOR ASSESSING RISK BENEFIT OF NEW DRUGS

The current system provides a framework for the assurance of evidence on safety and efficacy prior to the widespread marketing of new drugs in Canada. The process for licensing drugs embodies both the advantages and limitations of the available methodologies. The following sections highlight some of the limitations of these methodologies.

4.1 RCCT Sample Size Limitations

Sample sizes for randomized controlled clinical trials are designed to establish efficacy or effectiveness for the bulk of patients. These sample sizes are usually too small to predict rare events.[3] Furthermore, results of the trial may not generalize to patient populations other than those included in the trial.

4.2 RCCTs Usually Do Not Predict Long-Term Effects

Endpoints in most RCCTs are short or medium-term effects. Long-term follow-up is seldom included in an RCCT and it is therefore difficult to predict long-term drug effects. The importance of long-term follow-up for assessing drug safety is dramatically illustrated by the WHO trial on primary prevention of ischemic heart disease. In this trial, a healthy group of men with moderately increased serum cholesterol, treated with clofibrate to lower their serum cholesterol, had significantly decreased risk of non-fatal myocardial infarction compared to placebo controls. Follow-up (Committee of Principal Investigators, 1980) has demonstrated a persistent and significant increase in overall mortality for the clofibrate-treated group. Trials that do not have a large enough sample size or are not followed long enough to determine overall mortality effects can result in false negative conclusions.

[3] If, for example, a drug causes a serious adverse reaction in 1/1000 (0.1% or 0.001) of those who take it, the "rule of three" tells us that in order to be 95% confident of finding at least one adverse reaction we need to follow (1/00.001) X 3 or 3,000 patients. This is much larger than virtually any clinical trial. Although randomized trials are the only sure way to determine the efficacy of a drug, postmarketing surveillance is required to determine its safety (Sackett, Haynes, et al. 1980).

4.3 Statistical vs. Clinical Significance

A statistical significance probability level of $p = 0.05$ is arbitrary. A difference that is "statistically significant" may be clinically trivial. The question of "how much of an effect is effective?" needs to be defined prior to the study. Clinical importance is also a relative concept, however, and clinicians will often disagree on how much of a difference is important. Relative risk, odds ratios, absolute risk reduction, and number needed to be treated are clinical methods that seek to address this issue (Laupacis, Sackett and Roberts, 1988).

4.4 What is "Clinical Significance"?

Results may vary depending upon the measured outcome chosen. A recent study of surgery compared to medical care alone for patients at high risk of stroke was not able to determine any significant increase or decrease in the risk of stroke. However, functional status data collected during the surgery study showed that the surgery resulted in significant transient worsening of its functional status compared to medically treated patients (Haynes, Mukherjee, Sackett, et al. 1987). This example further emphasizes that the definition of clinically-relevant endpoints (in this case stroke vs. functional ability) critically determines whether or not results are significant.

4.5 Difficulty of Assessing Multiple Effects

It is difficult to assess multiple and often contradictory effects. For example, endpoints to determine pharmacologic versus therapeutic efficacy differ and are therefore difficult to integrate in terms of an overall risk-benefit ratio. This also applies to immediate, versus short-term versus long-term effects. Symptoms such as pain, discomfort or incapacitation are often deliberately omitted from assessment because the data are thought too "soft" for scientific attention, even though they may be more relevant for clinical goals. For example, most will agree that the treatment impact on the patient's family is non-trivial, yet this is rarely measured. A tastier diet or the frequency of injections, for example, clearly affect compliance and thus community effectiveness.

A common approach to assessing the risk-benefit ratio is to sum both the number of benefits and the number of side effects (Fineberg, 1971). The questions to address are:

a) What type of effects should be included?

b) How should these effects be defined and quantified?

c) How many effects should be quantified and with what degree of completeness?

The answers to these questions will vary depending on the nature of the drug treatment. There are few standardized measures and it is difficult to determine what level of risk is acceptable.

5. QUALITY OF LIFE IN THE ASSESSMENT OF PRESCRIPTION DRUGS

5.1 Definition of Health-Related Quality of Life

Quality of life QOL is a broad and multi-dimensional concept. The use of the term ranges from colloquial, to quite technical, to the narrowly defined meaning. Although the term "quality of life" virtually always refers to the satisfaction and enjoyment of human beings in a particular situation or environment, the specific meaning of quality of life in various applications and disciplines varies considerably.

One measure of QOL is real per capita income — a measure of the potential for consumption or the material standard of living. A broader concept of QOL is adopted in the social indicators literature in which economic dimensions are supplemented with information on local recreational and cultural amenities, educational opportunities, water and air quality, crime, and the nature and level of community services and activities.

Such measures are valid and useful for the purposes that they were developed. In the context of health care evaluation, these concepts of QOL are too broad. The group of measures that focus on the aspects of quality of life that are amenable to influence by health care interventions are referred to as measures of health-related quality of life.

Health care interventions have in the past been focused on reducing mortality. Society is, however, increasingly concerned with improving the quality as well as the longevity of life. Health-related quality of life (HRQOL) has emerged as a label for the measurement of a broad array of physical, social and emotional functions. For example, most QOL measures include the following dimensions or attributes:

a) sensory function (speech, sight, hearing),
b) mobility and physical function,
c) emotional function,
d) cognitive function,
e) self-care function regarding activities of daily living, and
f) the level of pain and discomfort.

Overall QOL is affected by a wide variety of factors other than health status. Socio-economic status is a prominent example. In the context of clinical studies, the assumption is that these other factors are balanced by the research design and therefore may be ignored for the purposes of making comparisons of the quality-of-life impact among alternative health care treatments. In the section that follows, the focus will be on the measurement of HRQOL. More specifically, the focus will be on the employment of these measures in the assessment of the risks and benefits of new drugs.

In a trial of a new drug, measurement of HRQOL has many dimensions. Benefits attributed to the drug must be weighed against the risk of drug-induced side effects. Other drug properties affect quality of life (and drug compliance/effectiveness) in addition to side effects. For example, in general, oral preparations are preferred to injected; self-administered hypodermic regimens are preferred to physician administered intravenous solution; tasty diets are preferred to bland; one dose of a long-acting drug is preferred to multiple doses of a shorter-acting drug; easier to swallow drugs are preferred; drugs that are easy to store, prepare, and transport are generally preferred over those that are not; and drugs that cost less are preferred to more expensive drugs.

Different approaches to the measurement of HRQOL are discussed in Section 5.2. One important approach is the utility approach. The characteristics of this approach are examined in Section 5.3. Measurement of utility is discussed in Section 5.4. The advantages and disadvantages of the use of utility-based QOL in the assessment of new drugs for licensing purposes are presented in Sections 6.1 and 6.2. Finally, in Section 6.3, an illustrative example of a utility-based QOL assessment of a new drug is presented. The drug is an oral gold compound (Auranofin) used in the treatment of arthritis patients.

5.2 Approaches to Measuring QOL

A wide variety of approaches to QOL assessment has emerged. Such a diversity is both inevitable and appropriate. Quality of life is a broad multi-dimensional concept. Different measures may tap different domains of quality of life. Investigators also have a variety of objectives and thus may require a battery of instruments and approaches for different research questions and situations. At present there is no gold standard for the measurement of quality of life, nor is one likely to develop.

There are a number of ways to classify quality-of-life measures. One classification scheme is presented in Table 2 (page 48). *Specific instruments* are developed for application in particular clinical situations. Typically, these measures are tailored to the quality-of-life burden of the illness (or health problem) and its treatment. By including only those elements that are most important in the particular clinical situation, these measures can be constructed to include a wide range of effects (for instance physical function, emotional function, characteristics that are of special concern to individual patients) without imposing a large burden on respondents. These measures usually have been constructed to maximize their responsiveness to change. A good example of a specific instrument is the Chronic Respiratory Disease Questionnaire developed by Guyatt et al. (1987).

Generic instruments are designed to be applied in a broad range of situations. These measures span a number of dimensions including physical function, social and emotional function, pain, self-care, and other activities. There are two major categories of generic measures, *Health Profiles and Health Utility Measures* (Guyatt, van Zanten, Feeny, and Patrick, 1988; and Feeny, Labelle, Torrance, in preparation):

a) In health profiles, scores are obtained for each relevant domain of health-related quality of life. The profile may or may not provide a single summary score. The Sickness Impact Profile (Bergner, 1981) is a good example of a health profile approach that provides a single summary score. The Nottingham Health Profile does not.

b) The second type of generic measure, the utility approach is discussed in the remaining sections. The following section provides a brief review of the foundations for the utility approach to assessing the quality of life.

5.3 Definition of "Utility"

Utility, a concept used in economics and decision analysis, refers to a level of satisfaction or enjoyment. In general, economists do not attempt to measure utility (the preferences of consumers for various consumption alternatives) directly. In the context of the analysis of behaviour in most market settings, preferences may be inferred from observing the behaviour of consumers. As a result, the discipline of economics has a well developed set of statistical tools for the analysis of data drawn from experience, but a much less extensive set of tools for the direct measurement of utility.

As opposed to economists, decision analysts regularly measure the preferences of decision makers. The measurements are founded directly on the axioms of utility theory first postulated by von Neumann and Morgenstern (1944). Although challenges to this approach continue to appear (Samuelson and Zeckhauser, 1988), these axioms have stood for over four decades as the dominant normative model for rational decision making when outcomes are uncertain. The classical method (gold standard) for measuring utilities is by use of the standard-gamble technique which comes directly from the basic axioms. Utilities so measured are properly called "von Neumann-Morgenstern utilities" to avoid confusion. It is the von Neumann and Morgenstern definition of utility that is used in this report.

5.4 Utility Measurement

A small family of approaches that are reliable and valid for the measurement of utility has been developed. The practical steps of the approach will be briefly outlined here; more detailed descriptions are found in: Torrance, 1986; Drummond, Stoddart, and Torrance, 1987; or Feeny and Torrance, 1989. Typically a number of health state scenarios are described, perhaps including the patient's subjectively-defined current health state. These scenarios may cover a wide range of dimensions, including physical function, sensory function, cognitive function, emotional well being, pain, self care, and any special features of the illness or its treatment (an example is provided in Table 3, page 49). Subjects are then asked to rank the health states, often with the assistance of a visual analog scale marked from 0 to 100, known as a feeling thermometer (see Figure 3, page 44). Scores on the feeling thermometer give the investigator a firm indication of the ordinal rankings of the health states (various treatment outcomes) and some information on the intensity of those preferences. Although the feeling thermometer may appear to give interval-scale values, evidence to date indicates that while it does provide a good indication of ordinal rankings, it does not directly provide cardinal utility scores.

The measurement of cardinal scores for the utility of health states is performed using one of two techniques: "standard gamble" or "time trade off". In the standard-gamble approach, the subject is offered a choice between two alternatives: choice A is to take the relevant health state with certainty, while choice B is to take a gamble on a treatment that has the possibility of a better or worse outcome. The application of this approach to the measurement of utility for a chronic state, called state A, is displayed in Figure 4 (page 45). Here choice B leads to perfect health with probability p and immediate death with probability $1-p$. The probability p is then varied until the subject is indifferent between Choices A and B. To see that there must be one and only one indifference probability, consider the case where state A is,

in the eyes of the subject, better than death but worse than perfect health. When p is 1.0 the subject will prefer choice B because that gives perfect health with certainty. When p is 0.0 the subject will prefer choice A because choice B now leads to immediate death with certainty. Somewhere between $p = 1.0$ and $p = 0.0$ there is a value of p, known as the indifference probability, at which the subject must switch.

The utility scale for health states is defined with a utility of 1.0 for perfect health and 0.0 for death. Then, the axioms of utility theory specify that the utility of state A is simply equal to the indifference probability. For example, if a subject is indifferent between choices A and B when $p = 0.75$, the utility of state A is 0.75. Note that the lower the indifference probability, the greater the risk of death the subject is willing to consider, and thus the lower the utility of the state A.

Because some respondents have experienced difficulty in understanding probabilities, an alternative technique, time trade off, has been developed. Although time trade off is not based directly on the axioms of utility theory, it has been found empirically to provide a close approximation (Torrance, 1976). The application of this approach to the measurement of utility for a chronic state is displayed in Figure 5 (page 46). In this technique the subject is also offered two alternatives. Choice A is the same as in the standard gamble: health state A for the rest of life, where the amount of remaining life is t years. Choice B is to take perfect health for a shorter period, x years, followed by death. Time x is varied until the subject is indifferent between the shorter period in perfect health and longer period in the less desirable state. The utility value of health state A is given by the ratio x/t. For example, if the subject is indifferent between 40 years of remaining life in health state A versus 30 years of remaining life being healthy, the utility value of health state A is 30/40 or 0.75.

Standard gamble, time trade off, and variations on these techniques provide reliable and valid methods for eliciting scores for health state utilities. The prior use of the feeling thermometer appears to assist respondents in their introspection concerning their preferences for the health states being evaluated. The health states to be evaluated may include states the respondent has experienced, is now experiencing or has never experienced (hypothetical states). In this regard, the utility approach differs from the other quality-of-life measures. The utility approach when it includes hypothetical states allows the investigator to obtain important information from all patients on how they think they would feel if they experienced some of the infrequent outcomes. Evidence to date shows that evaluations by persons experiencing the health state and by others for whom the health state is hypothetical usually do not differ substantially (Sackett and Torrance, 1978; Wolfson et al. 1982). If there is a

systematic difference, it is that person experiencing the health state rate that is marginally higher than those who have not experienced it, but this is not as yet conclusive. Thus the quality of life of rare or infrequent outcomes may be assessed along with that of the frequent outcomes by combining actual and hypothetical states in the set of outcomes for evaluation.

6. EXPLORING THE USE OF UTILITY-BASED QOL MEASUREMENT IN THE ASSESSMENT OF NEW DRUGS FOR LICENSING PURPOSES

The case for using utility measures of quality of life depends in part on the basic characteristics of QOL and advantages of the approach to QOL.

6.1 Advantages of Employing Utility Measures

The advantages of the utility approach to quality-of-life measurement include its generalizability, comprehensiveness, ability to integrate mortality and morbidity effects, ability to integrate risks and benefits, ability to elicit preferences both for hypothetical and actual situations, ability to represent multiple viewpoints, ability to use different types of evaluators for health states, its incorporation of time and risk preferences of subjects in their health state evaluations, its solid theoretical foundation, and its consistency with economic evaluation (see Table 4 page 50).

6.1.1 Generalizability

Scores derived from the utility approach are represented on a generalizable scale of ability to generalize from 0.0 (dead) to 1.0 (perfect health). The scale is completely general and can be used to measure the impact of all drugs no matter what the disease or type of patient.

6.1.2 Comprehensiveness

Utility measures are among the most comprehensive measures of quality of life. Health state descriptions typically include information on the levels of sensory, physical, socio-emotional, cognitive, and self-care functions, as well as the symptom or health problem, side effects, and the level of pain. Thus the evaluator of a particular health state integrates all the relevant characteristics of health-related quality of life into the utility-based measure of overall well-being.

6.1.3 Integration of Morbidity and Mortality Effects

The utility approach allows the analyst to combine mortality and morbidity effects into a single summary measure of health improvement. Normally, this is accomplished by combining the utility approach with the quality-adjusted life year (QALY) approach through the use of utilities as the quality adjustments (Weinstein

and Stason, 1977; Drummond, Stoddart and Torrance 1987; Torrance and Feeny 1990). It can also be accomplished directly within the utility approach by defining and measuring a utility function over the two major attributes, quantity of life and quality of life (Pliskin et al, 1980).

6.1.4 Capturing of Two Dimensions of Risk

Expected utility captures two dimensions of risk: (1) the "probability" of an outcome and (2) the "degree of preference" for that outcome. Other approaches to risk-benefit analysis are not able to capture both dimensions in one comprehensive number.

6.1.5 Integration of Risks and Benefits

The utility approach measures all impacts, positive and negative, of a new prescription drug in the same common units. In this way it provides the potential to integrate the risks and benefits, and to determine if the risk-benefit balance is favourable or unfavourable both within and across patients.

6.1.6 Preference Measurement for Hypothetical and Actual Health States

A sixth advantage is that both hypothetical and actual health states may be evaluated by the same person. Most quality-of-life measurement approaches and psychosocial tests require that the subject be experiencing the state when the assessment is performed. As a result, assessments may not be obtained for all relevant outcomes. The problem is particularly acute for states that are infrequent. In the utility approach, useful data can be obtained by evaluating the hypothetical states relative to actual ones for which scores are also obtained.

6.1.7 Representation of Multiple Viewpoints

A seventh characteristic of the utility approach is that a variety of evaluators for the same health states may be used including patients, relatives of patients, and members of the public. Comparisons among the evaluations can be important and informative. This tool allows for an assessment of a technology from several perspectives, such as the more narrow perspective of the patient and the broader societal viewpoint.

6.1.8 Incorporation of Time and Risk Preferences

An eighth characteristic of the utility approach is that it captures the rate of time preference (time trade off and standard gamble) and risk attitudes (standard gamble) of the respondent. Given that the uncertainty of the outcome is implicit in most clinical situations, the utility approach is an especially useful paradigm (pattern).

6.1.9 Rigorous Theoretical Foundation

An important strength of the utility approach is its rigorous, explicit, theoretical foundation based on the von Neumann and Morgenstern theory of utility.

6.1.10 Integration with Economic Evaluation Measures

Another advantage of the utility measurement approach is that it is constructed for integration with economic evaluation measures. This allows for comparisons of outcomes across diverse health care interventions through cost-utility analysis (CUA) and its variants. The conceptual foundations of utility assessment are consistent with the theoretical foundations of cost-benefit analysis in economics.

6.1.11 Integration with Clinical Trials

Utility-based quality of life measurements can be integrated into prospective clinical trials. Randomization and blinding provide for valid and unbiased measurement. Also, by combining clinical and quality-of-life investigations in one study, the results are all available in a timely fashion to affect clinical and health care policy.

6.1.12 Usefulness for Prescribing Physicians

Utility-based quality of life results from clinical testing for the licensing decision can be useful later by the prescribing physician. For example, the results may indicate, by patient type, the expected risk-benefit balance and its distribution. It should be possible to summarize and present this information in a way that later helps the prescribing physician and patient estimate the likely risk-benefit balance for him or her.

In sum, the utility approach allows for comprehensive assessments of health-related quality of life over a wide range of applications. It produces a single score on a generalizable scale from 0.0 (dead) to 1.0 (healthy) scale. Information on risk attitudes, time preference, and trade offs among the benefits of treatment and side effects are reflected in the utility scores. Utility analysis can also be readily combined with pecuniary (dollar) measures of costs and benefits of health care alternatives. Both measures can be incorporated into randomized clinical trials.

6.2 Disadvantages of Utility Approach to Measuring Quality of Life

The advantages of using the utility approach to measuring health-related quality of life do not come without accompanying disadvantages (Table 4, page 50):

a) First, although utility analysis measures are reliable and valid they are not highly precise. The standard error for individual scores is approximately 0.13 (Torrance, 1976).

b) Second, to obtain reliable measures, labour-intensive interviews conducted by trained professional interviewers are required.

c) Third, utility scores may not be readily understood or easily interpreted.

d) Fourth, although utility measures have been successfully obtained from a variety of types of interviewees, they cannot be universally applied. The complexity of the procedures appears to require native language ability in the language in which the materials are presented. There are age and cognitive ability restrictions as well.

e) Finally, while one of the major advantages of utility measurement is its generalizable scores based on subjective integration by the person, the approach does not reveal why a particular utility value may be high or low. Other quality-of-life measures may need to be included in the trial to allow for such determinations.

6.3 An Example of the Incorporation of Quality-of-Life Measurement in a Drug Trial

Quality-of-life measures were incorporated in a recent international multi-centred trial of a new pharmacological agent, Auranofin, for the treatment of arthritis (Bombardier et al, 1986). Quality-of-life measures were appropriate in this trial because improved quality of life, and not reduced mortality, was the relevant outcome. The quality of life for these patients would be affected not only by pain and other foibles of the illness, but also by the side effects of the treatment. Some of these side effects include dermatitis, stomatitis, thrombocytopenia, leucopenia, and proteinuria, diarrhea, loose stools and abdominal cramps. Serious gold toxicity is uncommon, and drug related deaths are rare in the treatment of rheumatoid arthritis.

The study compared Auranofin (3 mg twice daily) with a placebo as additional treatment for rheumatoid arthritis patients already stabilized on basic conservative therapy including rest, physical therapy, non-steroidal anti-inflammatory drugs, oral steroids and analgesics. Three hundred and three patients at fourteen centres were followed for six months in a randomized, double-blind design. A wide variety of outcome measures were employed (see Tables 5, page 51 and Table 6, page 52). The study provides one of the best head-on-head comparisons to date of alternative measures within a clinical trial. The utility instrument used in the trial was a "patient-utility measurement set" (PUMS) developed specifically for the study. The other outcome measures consisted of clinical measures, function-status measures, pain measures, global impression measures, and others.

The PUMS utility instrument is based on well established utility measurement techniques. It is designed to quantify, through a patient interview after treatment, the change in quality of life associated with treatment. The instrument consists of a user's manual and a questionnaire, supplemented by elaborate visual aids in the form of a card deck, with the whole package designed for ease of use by both the interviewer and the patient. It is designed to be applicable to adult patients of all ages, to neutralize potential framing biases, and to counter known measurement biases through a converging ping-pong strategy.

The PUMS instrument consisted of three parts:

- **Part 1** measured the change in quality of life between the baseline period before the trial and the time of interview. A change score was measured three ways: by rating scale, lottery, and modified time trade off. The rating scale was a variation on the feeling thermometer (Figure 3, page 44) adapted to measure a change score rather than an absolute score. Similarly, the lottery and the modified time trade off were change-score adaptations of the standard gamble (Figure 4, page 45) and time trade-off (Figure 5, page 46) respectively.
- **Part 2** measured the absolute quality of life at baseline and at the time of interview by conventional time trade off (Figure 5, page 46).
- **Part 3** measured "risk aversion"; more precisely, it measured the concavity/convexity of the utility function for additional healthy time. Part 3 used an adaptation of standard gamble (Figure 4, page 45).

The PUMS was administered to 239 patients in the trial (122 with Auranofin, 117 with a placebo) in fourteen different centres by interviewers at each centre. Average interview time was 32 minutes (range 8-82 minutes). Interviews were tape-recorded, 100 of the interviews were reviewed, and no significant difficulties were found in any parts of the instrument.

The PUMS results indicated that both Auranofin and placebo patients improved during the trial, but Auranofin patients improved more. Specifically, placebo patients improved from a quality of life of 0.629 to 0.728, an increase of 0.099 or 16%. Auranofin patients went from 0.571 to 0.781 for an increase of 0.209 or 37%. The difference was significant at $p = 0.002$.

A wide array of outcome measures were used in the trial and not all were equally responsive (see Tables 5, page 51 — and Table 6, page 52). Auranofin was found to be superior to placebo by a composite of the clinical outcome measures ($p = 0.003$), by a composite of the function-status measures ($p = 0.001$), by a composite of the global impression measures ($p = 0.007$) and by Bush's Quality of Well Being ($p = 0.005$). A composite pain measure showed improvement ($p = 0.021$) but did not attain the trial design level of significance of $p = 0.0125$. A conventional standard gamble against health and death at the end of treatment gave numerical results similar to the PUMS (utility of Auranofin 0.77 and of placebo 0.70 compared to the PUMS values of 0.78 and 0.73), but did not quite reach statistical significance ($p = 0.07$), probably because it was not designed to measure change as was the PUMS.

Virtually no difference was detectable at all by the willingness-to-pay method, a relatively new approach that is known to have serious measurement difficulties (Klarman, 1982; Thompson et al, 1982; Thompson et al, 1984). Essentially the willingness-to-pay method attempts to determine the dollar value that a person would be willing to pay to rid themselves of the condition. Although Table 6, (page 52), taken from the original article (Bombardier et al, 1986), classifies willingness to pay as a utility measure, that is not strictly correct. Willingness to pay gives a dollar measure, not a utility measure; however, this particular dollar measure is designed to replace the need for utility measures in economic evaluations.

Similarly, no difference was shown by the NIMH Depression Scale, while only a modest non-significant difference was found by the RAND General Health Perceptions Questionnaire. Neither of these are utility measures.

The PUMS utility instrument proved to be one of the most responsive outcome instruments in the trial, as indicated by the level of statistical significance attained. Indeed, it was only exceeded in responsiveness by two measures, the composite function-status score and the arthritis categorical scale. Moreover, composite scores have attenuated statistical power compared to scores of individual instruments. For example, each of the three individual components of the function-status composite was less responsive than the PUMS. PUMS, unlike these other measures, also provides an assessment of the global impact on quality of life.

Thus, comparing the 25 individual outcome measures head-to-head in this trial, the responsiveness of PUMS was only exceeded by that of the arthritis categorical scale. In the latter the patients are basically indicating that their arthritis has improved. In the PUMS they are translating that improvement into an overall impact on their health-related quality of life.

Note: The PUMS instrument is study-specific and cannot be used directly in other studies without modification. Nonetheless, its success in this study suggests that similar instruments can be developed for other clinical trials, and that the utility approach is appropriate whenever a comprehensive evaluation of outcome is desired.

One of the major strengths of the PUMS instrument is its simplicity of use. The essence of the instrument is a card deck that is almost self-guiding, supplemented by a precise set of interview instructions. With careful training and quality control, research assistants at fourteen different centres in two countries were able to use the instrument in a consistent manner without problems.

The Auranofin example illustrates that the utility measurement approach can be viably incorporated into clinical trials and used as a measure of the quality of life of outcomes. Experience indicates that when study-specific utility instruments are carefully developed and deployed, they are reliable, valid, and responsive.

7. SUMMARY AND CONCLUSIONS

Reduction in the burden of illness and improvement in the quality of life are among the treatment goals of many new pharmaceutical agents. In addition, many if not most drug therapies involve side effects and/or adverse reactions that impact on the quality of life of persons using that medication. Given that quality of life is a key outcome, it is important and desirable to measure it directly in drug trials.

The recent and rapid development of improved methodologies for measuring health-related quality of life now makes it possible to measure quality-of-life effects directly in clinical trials. A number of specific instruments have been used successfully in clinical trials of drugs for a wide range of conditions, including chronic respiratory disease, chemotherapy for the treatment of breast cancer, arthritis, and hypertension. A number of generic instruments such as the Sickness Impact Profile have also been successfully deployed in a variety of clinical settings. Finally, a few trials have successfully included utility measures of health-related quality of life. In general in these applications, quality-of-life measures complemented and extended traditional clinical measures of outcome. Quality-of-life measures both added information to that available from other measures and had the further advantage of measuring directly the outcome that is often of fundamental importance: the patient's quality of life. Although additional methodological development is still needed and a gold standard is lacking, both the pharmaceutical industry and the regulatory authorities such as the HPB should develop further experience and expertise with these methodologies. It is now reasonable to consider evidence based on the utilization of these tools in new drug submissions. Health-related quality of life is a valid measure of outcome.

The use of quality-of-life measures adds important information to that previously available on treatment effects. The development of reliable tools to objectively measure subjective phenomenon such as changes in quality of life adds an important dimension to risk-benefit analysis. Disease-specific quality-of-life instruments and generic measures such as health profiles provide information on the changes in health status perceived by patients and the importance patients assign to those changes. These measurement instruments also often record the quality-of-life impacts of side effects. A number of these instruments have proved useful in assessing a small but diverse number of drugs.

The utilization of any reliable, valid, and responsive measure of health-related quality of life adds useful information to risk-benefit analysis. One approach, however, provides a more comprehensive and rigorous framework within which to assess the trade offs among both positive and negative treatment effects. The utility approach generates a single summary score that reflects the net impact on quality of life. Patients subjectively weigh, according to their preferences, the value of improvements in health status versus the burden of side effects resulting from the drug therapy. By combining hypothetical and actual health states, the repercussions of infrequent but important side effects may also be represented in the same evaluation. In addition, the utility approach integrates mortality and morbidity effects, thus providing a very comprehensive framework with which to assess the effects of drug therapy. The utility approach is highly relevant and attractive as a means of extending traditional risk-benefit analysis.

8. FINALE

Methodologies for measuring health-related quality of life, including the utility approach, are now sufficiently well developed that it is scientifically appropriate to use them modestly in drug trials. Quality-of-life measures add significant and useful information to that which is traditionally submitted for drug licensing. The utility approach to quality-of-life measurement is particularly powerful, appealing, and deserving of consideration.

9. REFERENCES

Altshuler, B. 1976. A Bayesian Approach to Assessing Population Risks from Environmental Carcinogens. *Proceedings SIMS Conference on Environmental Health.* Alta, Utah.

Banta H.D., C.J. Behney and J.S. Willems. 1981. *Toward Rational Technology in Health Care.* New York: Springer Publishing Company, Inc. p. 93.

Bergner, M., R.A. Bobbitt, W.B. Carter and B.S. Gilson. 1981. The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. *Medical Care.* 19: 787–805.

Bombardier, C., J. Ware, IL. Russell, M. Larson, A. Chalmers and J.L. Read. 1986. Auranofin Therapy and Quality of life in Patients with Rheumatoid Arthritis. *Am. J. Med.* 81: 565–578.

Buxton, M.J., M. F. Drummond. 1988. Quality of Life Measurement and the Regulation of Medicines. (unpublished draft).

Committee of Principal Investigators. 1980. W.H.O. Cooperative Trial on Primary Prevention of Ischemic Heart Disease Using Clofibrate to Lower Serum Cholesterol: Mortality Follow-up. *Lancet.* August 23, 1980: 379–385.

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton. 1984. How to Read Clinical Journals: VII. To Understand an Economic Evaluation (Part A). *Can. Med. Assoc. J.* 130: 1428–1433.

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton. 1981. How to Read Clinical Journals: IV. To Determine Etiology or Causation. *Can. Med. Assoc. J.* 124: 985–990.

Drummond M.F., G.L. Stoddart and G.W. Torrance. 1987. *Methods for the Economic Evaluation of Health Care Programs.* Oxford: Oxford University Press.

Feeny, D. 1988. Can We Use Quality-of-Life Measures for Economic Evaluation? *Quality of Life in Cardiovascular Care* (forthcoming).

Feeny, Barr, Furlong, Torrance, Rosenbaum, Weitzman, and Horsman (in preparation). Economic Evaluation of Treatment for Childhood Cancer: Costs, Benefits, and the Quality of Life.

Feeny, D., G. Guyatt and P. Tugwell, ed. 1986. *Health Care Technology: Effectiveness, Efficiency and Public Policy.* The Institute for Research on Public Policy. Montreal.

Feeny, D., R. Labelle and G.W. Torrance. 1990 (in press). Chapter 6: Integrating Economic Evaluations and Quality-of-Life Assessments. In *Quality of Life Studies*. Edited by Bert Spilker. Raven Press, New York.

Feeny, D., and G.W. Torrance. 1989. Incorporating Utility-Based Quality-of-Life Assessment Measures in Clinical Trials: Two Examples. *Medical Care*. 27(3):S190–S204.

Feinstein, A. R. Clinical Biostatistics. IX. How Do We Measure "Safety and Efficacy"? *Clin. Pharm. S. Ther.* 12(3): 544–558.

Fineberg, H.V. 1980. Decision Trees: Construction, Uses and Limits. *Bull. Cancer*. 67: 395–404.

Gilman, A., L.S. Goodman, T.W. Rall and M. Freid, ed. 1985. *The Pharmacological Basis of Therapeutics* (7th ed.). Macmillan Publishing Co., London.

Guyatt, G.H., M. Townsend, L.B. Berman and S.O. Pugsley. 1987. Quality of Life in Patients with Chronic Airflow Limitations. *Br. J. Dis. Chest*. 81: 45–54.

Guyatt, van Zanten, Feeny, and Patrick (in preparation). Measuring Quality of Life in Clinical Trials. manuscript, 1988.

Haynes, R.B., J. Mukherjee, D.L. Sackett, et al. 1987. Functional Status Following Medical or Surgical Treatment for Cerebral Ischemia. *J. Am. Med. Assoc.* 257(15): 2043–2046.

Jones, D.R., and R.L. Akehurst. 1980. Risk Assessment: An Outline. *The Institute of Mathematics and its Applications*. 16: 252–258.

Klarman, H.E. 1982. The Road to Cost-effectiveness Analysis. *Milbank Memorial Fund Quarterly / Health and Society*. 60(4): 585–603.

Knight, F. 1921. *Risk, Uncertainty, and Profit*. Houghton Miffin Co., Boston.

Laupacis, A., D.L.Sackett, R.S. Roberts. 1988. An Assessment of Clinically Useful Measures of the Consequences of Treatment. *N. Engl. J. Med.* 318(26): 1728–1733.

Laurence, D.R., A.E.M. McLean and M. Weatherall, ed. 1984. *Safety Testing of New Drugs: Laboratory Predictions and Clinical Performance*. London Academic Press, London.

Lindgren, B.W. 1968. *Statistical Theory*, 2nd ed. Macmillan, New York.

Linstone, H.A. and M. Twoff, ed. 1975. *The Delphi Method*. Addison Wesley, Reading, Mass.

Lowrance, W. 1976. *Of Acceptable Risk*. William Kaufmann Inc. Los Altos, Cal.

Mailhot, R. 1986. The Canadian Drug Regulatory Process. *J. Clin. Pharmacol.* 26: 232–239.

Napke, E. 1986. Canada. In *Monitoring for Drug Safety*, 2nd ed. Edited by W.H.W. Inmann. MTP Press Ltd., Boston. pp. 49–69.

O'Brien, B. 1986. *"What Are My Chances Doctor?" — A Review of Clinical Risks*. Office of Health Economics, London.

Office of Technology Assessment. 1983. *The Impact Of Randomized Clinical Trials on Health Policy and Medical Practice: Background Paper*. U.S. Congress, Office of Technology Assessment, OTA-BP-H-22. Washington, D.C.

Office of Technology Assessment. 1978. *Assessing the Efficacy and Safety of Medical Technologies*. U.S. Government Printing Office. Washington, D.C.

Office of Technology Assessment. 1976. *Cancer Testing Technology and Sacharin*. Publication No. OTA-H-34. Washington, D.C.

Pliskin, J.S., D.S. Shepard and M.C. Weinstein. 1980. Utility Functions for Life Years and Health Status. *Operations Research*. 28(1): 206–224.

Royal Society, The. 1983. *Risk Assessment: Report of a Royal Society Working Party*. Royal Society, London.

Sackett, D.L., and M. Gent. 1979. Controversy in Counting and Attributing Events in Clinical Trials. *N. Engl. J. Med.* 131: 1410–1412.

Sackett, D.L., R.B. Haynes, M. Gent and M.W. Taylor. 1980. Compliance. In *Monitoring for Drug Safety*. Edited by W.H.W. Inman. MTP, Lancaster.

Sackett, D.L., R.B. Haynes and P. Tugwell. 1985. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Little, Brown and Company, Toronto.

Sackett, D.L., and G.W. Torrance. 1978. The Utility of Different Health States as Perceived by the General Public. *J. Chron. Dis.* 31(11): 697–704.

Samuelson, W. and R. Zeckhauser. 1988. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty*. 1(1): 7–59.

Spetzler, C.S., and C.A.S.S. von Holstein. 1975. Probability Encoding in Decision Analysis. *Management Science*. 22: 340–358.

Spilker, B. 1984. *Guide to Clinical Studies and Developing Protocols*. Raven Press, New York.

Teeling-Smith, G., ed. 1983. *Measuring the Social Benefits of Medicine*. Office of Health Economics. White Crescent Press Ltd., Luton, U.K. p. 5.

Thompson, M.S., L.R. Read, J. Leighton, and M. Liang. 1982. Feasibility of Willingness-to-pay Measurement in Chronic Arthritis. *Med. Decision Making*, 4(2): 195–215.

Thompson, M.S., L.R. Read and M. Liang. Willingness-to-pay Concepts for Societal Decisions in Health. In *Values and Long-Term Care*. Edited by Kane, R. and Kane, R. D.C. Health Publisher. Lexington, Mass., pp 103–.

Torrance, G.W. and Feeny, D. 1989. Utilities and Quality-adjusted Life-years. *Int'l J. of Assessment in Health Care* (in press).

Torrance, G.W. 1986. Measurement of Health State Utilities for Economic Appraisal: A Review. *J. Health Economics*. 5: 1–30.

Torrance, G.W. 1976. Social Preferences for Health States: An Empirical Evaluation of Three Measurement Techniques. *Socio-Economic Planning Science*. 10(3): 129–136.

Treasury Board of Canada. 1979. *Administrative Policy Manual*. Chapter 490, Appendix E. Supply and Services Canada, Ottawa.

Tugwell P., K.J. Bennett, D. Feeny, G. Guyatt and R.B. Haynes. 1986. A Framework for the Evaluation of Technology: The Technology Assessment Iterative Loop. In *Health Care Technology: Effectiveness, Efficiency and Public Policy*. Edited by D. Feeny, G. Guyatt and P. Tugwell, ed. Institute for Research on Public Policy, Montreal. pp. 41–56.

Tugwell P., Bennett K.J., D.L. Sackett and R.B.Haynes. 1985. The Measurement Iterative Loop: A Framework for the Critical Appraisal of Need, Benefits and Costs of Health Interventions. *J. Chron. Dis.* 38(4): 339–351.

von Neumann, J. and O. Morgenstern. 1944. *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, N.J.

Walker, S.R. and A.W. Asscher, ed. 1987. *Medicines and Risk/Benefit Decisions*. MTP Press, Lancaster.

Webster's New World Dictionary, 2nd ed. 1980. Simon and Schuster, New York.

Weinstein, M.C. and H.V. Fineberg. 1980. *Clinical Decision Analysis*. W.B. Saunders Company, Philadelphia.

Weinstein, M.C. and W.B. Stason. 1977. Foundations of Cost-effectiveness Analysis for Health and Medical Practices. *N. Engl. J. Med.* 296(13): 716–721.

Wolfson, A.D., A.J. Sinclair, C. Bombardier and A. McGeer. 1982. Preference Measurements for Functional Status in Stroke Patients: Inter-rater and Inter-technique Comparisons. In *Values and Long-Term Care*. Edited by R. Kane and R. Kane. D.C. Health Publisher. Lexington, Mass.

Zbinden, G. 1982. Current Trends in Safety Testing and Toxicological Research. *Naturwissenschaften*. 69: 255–259.

Appendix A

Selected Major Recent Conferences on the Measurement of Health-related Quality of Life

CONFERENCE	DATES	SPONSOR(S)	PLACE
Epidemiology: An Essential Tool for Drug Development	Nov. 14–15, 1988	Drug Information Assoc.	Charleston, South Carolina
Advances in Health Assessment Conference	July 13–15, 1988	Institute of Medicine, Henry J. Kaiser Family Foundation	Menlo Park, California
Workshop on Quality of Life and Cardiovascular Disease	June 7–10, 1988	National Heart, Lung and Blood Institute, Bowman Gray School of Medicine	Winston-Salem, North Carolina
Cost Benefit Assessment of Pharmaceutical Products	March 14–16, 1988	Drug Information Assoc.	Hilton Head Island, South Carolina
Measuring Quality of Life with Pharmaceuticals: Fundamentals and Techniques	December 10–11, 1987	Pharmaceutical Manufacturers Assoc.	Baltimore, Maryland
Issues of Quality of Life in Transplantation and Oncology	May 20, 1987	Netherlands Health Council, WHO Amsterdam Collaborating Centre on Quality of Life*, Netherlands Cancer Foundation	Netherlands Cancer Institute, Amsterdam
Centre for Medicines Research Workshop	March 3, 1987	Centre for Medicines Research	CIBA Foundation, London
Measuring Quality of Life and Functional Status in Clinical Epidemiological Research	April 9–13, 1986	State of Science 1986	Sintra, Portugal
Advances in Health Assessment Conference	February 19–21, 1986	Henry J. Kaiser Family Foundation, the RAND Corporation	Palm Springs, California

* The "Amsterdam Collaborating Centre for Research and Training in Methods of Assessment of Quality of Life in Relation to Health Care" pursues a range of activities related to the development and implementation of research programs into the quality of life of chronic disease patients.

Appendix B

Major Bibliographic Resources

Clearinghouse on Health Indexes

U.S. Department of Health and Human Services
Public Health Service
National Centre for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

This special information centre was established in 1973 to promote the development of techniques that have become to be known as quality of life.

The clearinghouse has two main functions. One is the publication of the *Bibliography on Health Indexes*, which can be ordered from the U.S. Government printing office for approximately $2.50 per issue. The other main function is to serve as an information channel between the developers of quality of life assessments and clinical investigators; this exchange is usually handled informally, usually in a telephone conversation between clearinghouse staff and prospective investigators.

Annotated Bibliography for Risk Perception and Risk Communication

Centre for Research on Judgment and Policy
University of Colorado at Boulder
Attn: Doreen Victor
Campus Box 344
University of Colorado
Boulder, CO 80309–0344

This bibliography is divided into three major categories: risk perception, risk communication, and health related risk communication. Each section contains annotations of a broad range of journal articles, chapters, books, working papers, and meeting abstracts related to its heading. This bibliography includes references from the early 1970s through February 1988 (US: $25.00).

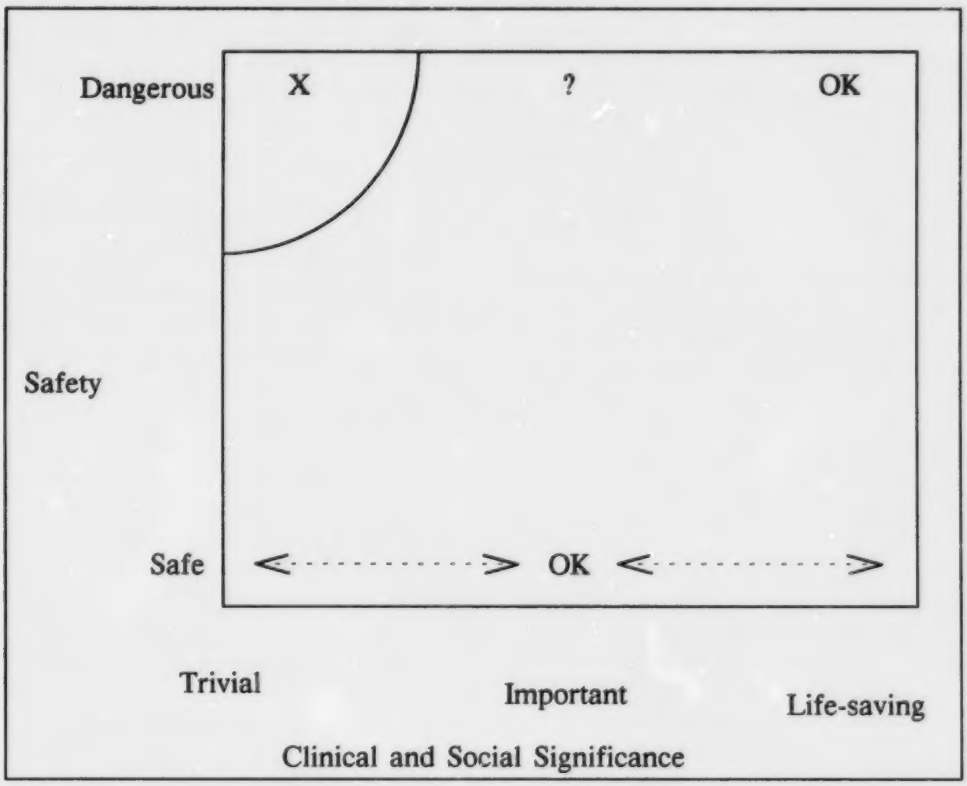

Figure 1: **The benefit-risk matrix***

* Source: Teeling-Smith, 1983.

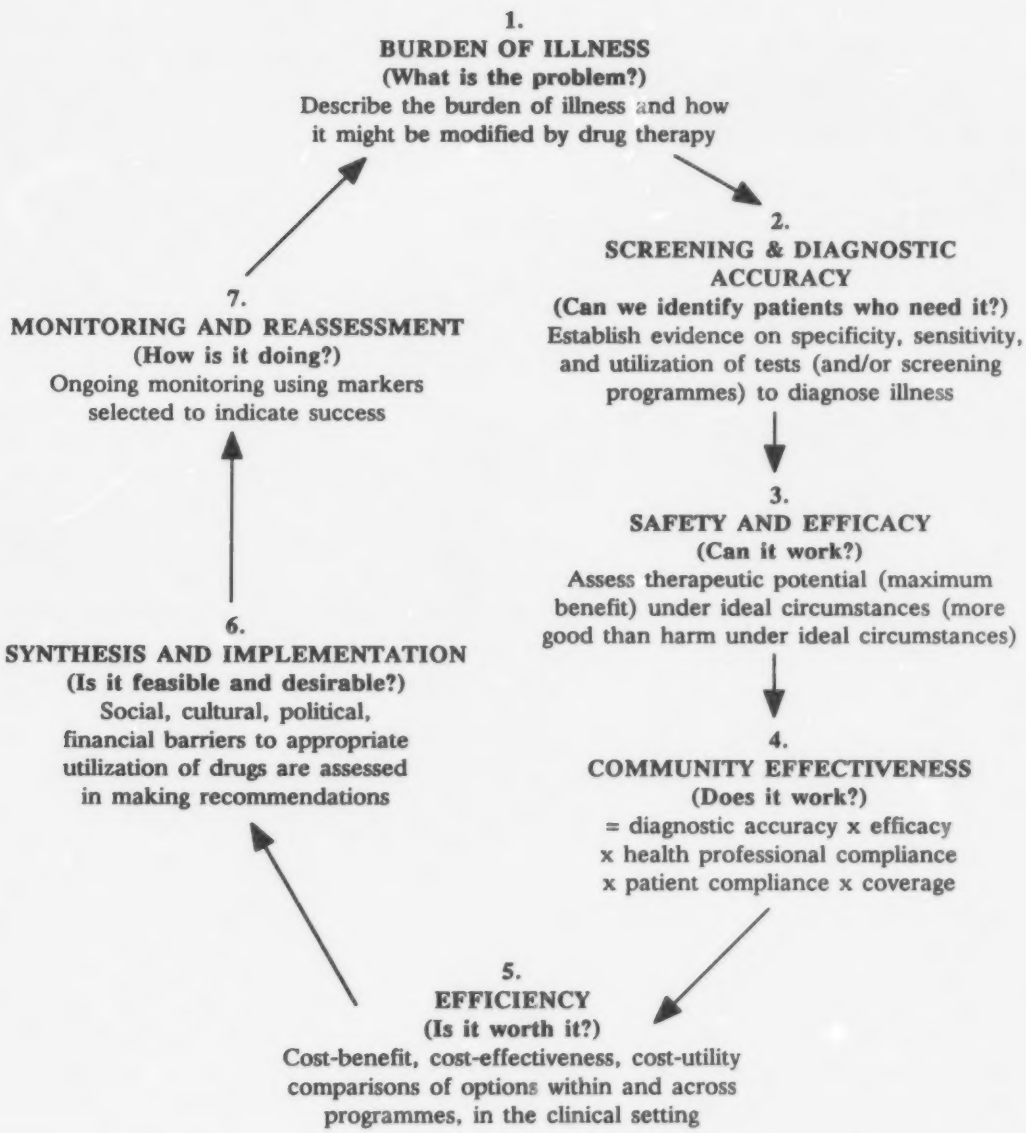

Figure 2: **The drug assessment iterative loop***

* Source: Modified from Tugwell et. al. (1985, 1986)

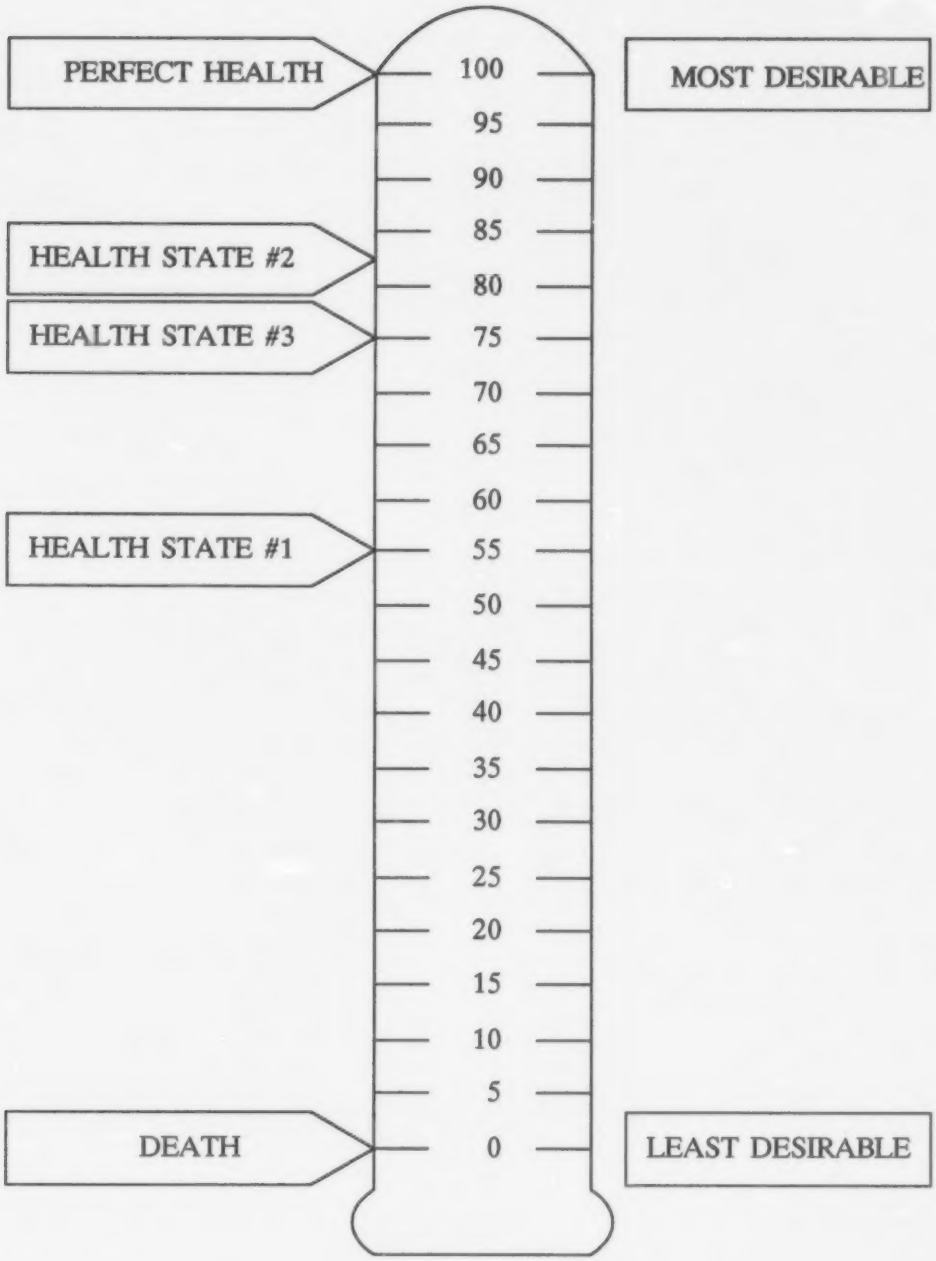

Figure 3: **Feeling thermometer approach for ordinal ranking**

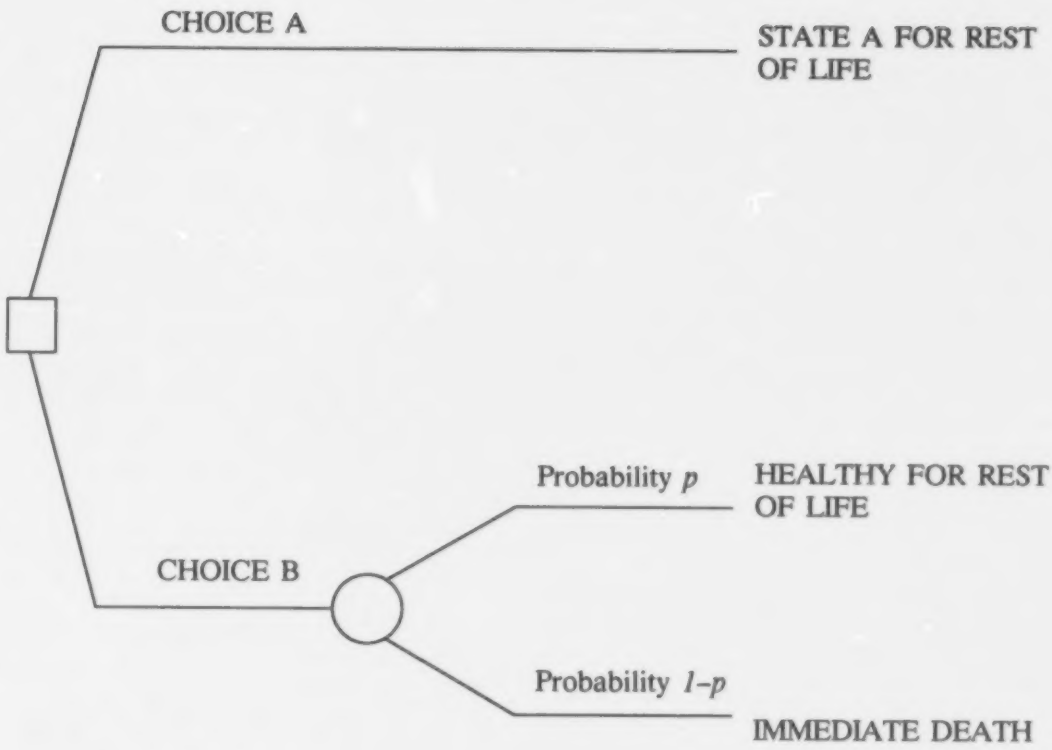

Figure 4: **Standard gamble approach for eliciting utility value for a chronic state**

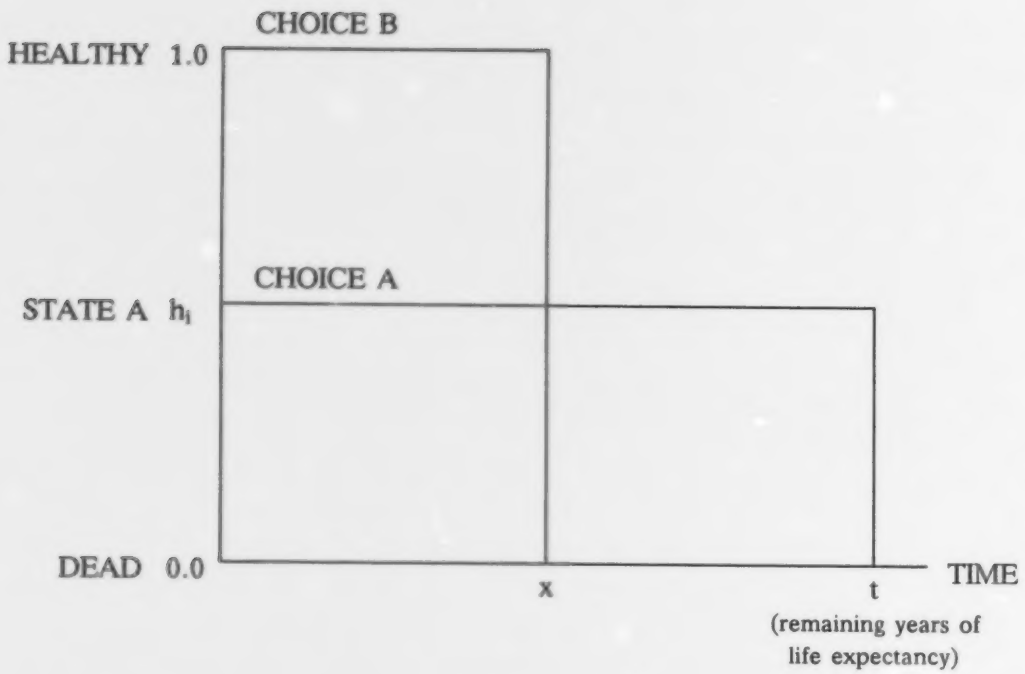

Figure 5: **Time trade off approach for eliciting utility value for a chronic state**

Table 1: **Selected Definitions of "Efficacy"***

Source	Term Defined	Definition
Federal Food, Drug, and Cosmetic Act	Efficacy, Effectiveness (used interchangeably)	A drug is effective if it has "the effect it purports or is represented to have under the conditions or use prescribed, recommended, or suggested in the proposed labelling thereof"
Cochrane in Effectiveness and Efficiency	Efficacy, Effectiveness (used interchangeably)	"The effect of a particular medical action in altering the natural history of a disease for the better"
World Health Organization	Efficacy	Benefit or utility to the individual of the service, treatment regimen, drug, preventive or control measure advocated or applied
Discursive Dictionary of Health Care	Efficacy (as a variant of effectiveness)	"The degree to which diagnostic, preventive, therapeutic, or other action or actions (undertaken under ideal circumstances) achieves the desired result"
Office of Technology Assessment	Efficacy	The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use
Department of Clinical Epidemiology and Biostatistics, McMaster University	Efficacy	Does the health procedure, service or program do more good than harm to people who fully comply with the associated recommendations or treatments?

* Adapted from:

(1) Office of Technology Assessment. *Assessing the Efficacy and Safety of Medical Technologies.* Washington, D.C.: U.S. Government Printing Office, 1978

(2) Department of Clinical Epidemiology and Biostatistics: How to read clinical journals: VII. To understand an economic evaluation (part A). *Can. Med. Assoc. J.* Vol. 130, June 1, 1983.

Table 2: **Taxonomy of Quality-of-Life Assessment Measures***

Category	Examples
A. SPECIFIC INSTRUMENTS	Chronic Respiratory Questionnaire Arthritis Categorical Scale Arthritis Impact Measurement Set
B. GENERIC INSTRUMENTS:	
B1. HEALTH PROFILES	Sickness Impact Profile Nottingham Health Profile
B2. UTILITY MEASURES	Standard Gamble Time Trade Off

* Source: Guyatt, van Zanten, Feeny, and Patrick (in preparation). *Measuring Quality of Life in Clinical Trials*. manuscript, 1988.

Table 3: **Example of Health-State Description***

Able to see, hear and speak normally for age.

Requires the help of another person to walk or get around; and requires mechanical equipment as well.

Occasionally fretful, angry, irritable, anxious, depressed or suffering "night terrors".

Learns and remembers school work normally for age.

Eats, bathes, dresses and uses toilet normally for age.

Free of pain and discomfort.

* Source: Feeny, Barr, Furlong, Torrance, Rosenbaum, Weitzman, and Horsman (in preparation). *Economic Evaluation of Treatment for Childhood Cancer: Costs, Benefits, and the Quality of Life.*

Table 4: **The Advantages and Disadvantages of Utility Measures***

A.	Advantages
1.	Generalizability
2.	Comprehensiveness
3.	Integration of Morbidity and Mortality Effects
4.	Captures Two Dimensions of Risk
5.	Preference Measurement for Hypothetical and Actual Health-states
6.	Representation of Multiple Viewpoints
7.	Incorporation of Time and Risk Preferences
8.	Rigorous Theoretical Foundation
9.	Integration with Economic Evaluation Measures
10.	Integration with Clinical Trials
B.	Disadvantages
1.	Utility Measures Not Highly Precise
2.	Require Labour Intensive Interviews
3.	Utility Scores Not Easily Understood
4.	Utility Measures Not Universally Applicable
5.	Does Not Reveal Why a Utility Value May Be High or Low

* Source: Based on Feeny and Torrance, 1989.

Table 5: **Changes in Clinical and Health Status Measures and Composite Scores between Baseline and Sixth Month By Treatment Group (n = 303)***

	Possible Range (worst to best)	Placebo (n = 149)	Auranofin (n = 154)	Treatment Effect (p Value)**
Clinical Measures				
Number of tender joints		−4.5 ± 0.9***	−7.3 ± 1.0	0.01
Number of swollen joints		−3.6 ± 0.7	−5.5 ± 0.7	0.01
50-foot walk time (sec.)		+0.61 ± 0.66	−0.74 ± 0.40	0.11
Duration morning stiff.(min)		− 12 ± 13	−28 ± 12	0.10
Grip strength adjusted (mm Hg)		−2 ± 8	+13 ± 8	0.02
Composite score		0.16 ± 0.05	0.35 ± 0.05	0.003
Function Measures				
Health Assessment Questionnaire	3–0	−0.17 ± 0.04	−0.31 ± 0.05	0.01
Keitel Assessment	98–0	+1.7 ± 1.0	−1.5 ± 0.9	0.003
Quality of Well-Being Question.	0–1	−0.001 ± 0.007	+0.023 ± 0.007	0.005
Composite score		0.05 ± 0.05	0.28 ± 0.05	0.001
Pain Measures				
McGill Pain Questionnaire	78–0	−5.6 ± 1.0	−8.0 ± 1.1	0.02
Pain Ladder Scale (6 day mean)	0–10	+0.61 ± 0.13	+0.96 ± 0.18	0.09
10-Centimetre Pain Line	10–0	−0.91 ± 0.18	−1.44 ± 0.17	0.01
Composite score		0.48 ± 0.07	0.74 ± 0.08	0.021
Global Impression Measures				
Arthritis:				
Categorical Scale	1–5	+0.31 ± 0.07	+0.65 ± 0.08	<0.001
Ladder Scale (6 day mean)	0–10	+0.59 ± 0.14	+0.98 ± 0.18	0.11
Overall health:				
Ladder Scale, current	0–10	+0.39 ± 0.14	+0.65 ± 0.15	0.19
Ladder Scale, 6-day mean	0–10	+0.35 ± 0.12	+0.87 ± 0.15	0.007
Rand Current Health Assessment	9–45	+0.51 ± 0.47	+1.82 ± 0.49	0.01
10-Centimetre Line, by patient	0–10	+0.51 ± 0.13	+0.89 ± 0.16	0.1
10-Centimetre Line, by physician	0–10	+0.46 ± 0.12	+0.58 ± 0.13	0.2
Composite score		0.27 ± 0.05	0.50 ± 0.07	0.007

* Source: Bombardier, Ware, Russell, et al. 1986.

** Significance level of the treatment effect determined by analysis of covariance for baseline values of age, sex, clinic, functional class, and the variable tested.

*** Mean change ± SEM.

Table 6: Results of Utility and Other Measures*

	Possible Range (worst (to best)	Placebo		Auranofin		Treatment Effect (p Value)
		Value or Change*	No. of Patients	Value or Change**	No. of Patients	
Utility measures						
Patient Utility Measurement Set (PUMS) (derived change score)[1]	*0–100*	9.9 ± 2.2	*117*	20.9 ± 2.6	*122*	*0.002+*
Standard Gamble (percent chance of death)[2]	*100–0*	30 ± 3	*124*	23 ± 2	*119*	*0.07+*
Willingness-to-Pay (percent of income)[2]	*100–0*	23 ± 1	*105*	21 ± 1	*112*	*0.79+*
Other Measures						
NIMH Depression Questionnaire[3]	*60–0*	*−4.1 ± 0.8*	*149*	*−3.3 ± 1.1*	*154*	*0.54++*
Rand General Health Perceptions Questionnaire[3]	*0–110*	*−0.07 ± 0.75*	*149*	*+0.52± 0.79*	*154*	*0.32++*
Clinical and functional measures not included in composites						
Erythrocyte sedimentation rate (mm/hour)[3]		*+0.57 ± 1.68*	*149*	*−4.31± 1.6*	*154*	*0.02++*
Toronto Activities of Daily Living Quest.(perceived change score)[4]	*4–1*	*0.00 ± 0.006*	*149*	*0.02 ± 0.007*	*154*	*0.02+*

* Source: Bombardier, Ware, Russell et al. 1986.

** Mean ± SEM.

\+ Significance level of treatment effect determined by analysis of covariance adjusting for baseline values of age, sex, clinic, and functional class but not the variable tested.

++ Significance level of treatment effect determined by analysis of covariance adjusting for baseline values of age, sex, clinic, functional class, and the variable tested.

1 No value obtained at baseline: derived retrospectively at fifth month from patient's comparisons of baseline and current states.

2 No value obtained at baseline: not a change score.

3 Values obtained at baseline and sixth month.

4 No value obtained at baseline; recorded retrospectively at sixth month.

Analyses risques-avantages et qualité de vie des médicaments prescrits

Santé et Bien-être social Canada
Health and Welfare Canada

Canada

ANALYSES RISQUES-AVANTAGES ET QUALITÉ DE VIE DES MÉDICAMENTS PRESCRITS

George Torrance, Ph.D.[1,2,3]
Barbara Markham, M.B.A.[1,2]
David Feeny, Ph.D.[1,2,4]
David Rosenbloom, Pharm.D.[5,6]
Amiram Gafni, Ph.D.[1,2]

1 Centre d'économie sanitaire et d'analyse de la politique de la santé, Université McMaster.
2 Département d'épidémiologie clinique et de biostatistique, Université McMaster.
3 Faculté d'administration, Université McMaster.
4 Département d'économique, Université McMaster.
5 Département de médecine, Université McMaster.
6 Département de services pharmaceutiques, hôpitaux Chedoke-McMaster.

Santé et Bien-être social Canada
Mai 1990

DD-90-7

© Minister of Supply and Services Canada

1990

Catalogue No. H42-2/24-1990

ISBN 0-660-55555-7

AVANT-PROPOS

Un examen des articles publiés sur les méthodes d'analyse des risques-avantages et de la qualité de vie associés aux médicaments de prescription a été effectué. Il ressort de cette étude que l'utilisation des techniques de mesure de la qualité de vie est encouragée dans les essais cliniques de certaines drogues nouvelles. En particulier, la mesure de l'utilité est reconnue comme une technique puissante pour la détermination du bilan des risques et des avantages que présentent de nombreuses drogues nouvelles.

TABLE DES MATIÈRES

Page

1. RÉSUMÉ 1

2. INTRODUCTION 4

3. ANALYSE RISQUES-AVANTAGES DES DROGUES NOUVELLES 5

3.1 Boucle itérative d'évaluation d'un médicament 5
3.1.1 Portée du rapport 9

3.2 Risques, avantages et analyses risques-avantages : définitions 9
3.2.1 Risque (innocuité) 9
3.2.2 Avantages (efficacité théorique/efficacité pratique/rentabilité) 10
3.2.3 Analyse risques-avantages 11

3.3 Méthodes de mesure des risques-avantages 12
3.3.1 Identification 12
3.3.2 Quantification 12
3.3.3 Évaluation 13

3.4 Autorisation des nouveaux produits de prescription 15
3.4.1 Revue des conditions requises pour l'autorisation des médicaments au Canada 15
3.4.2 Approches actuelles utilisées par les compagnies pharmaceutiques ... 16

4. LIMITES DU SYSTÈME ACTUEL D'ÉVALUATION DU RAPPORT RISQUES-AVANTAGES DES DROGUES NOUVELLES 18

4.1 Essais cliniques randomisés — limites dues à la taille des échantillons 18

4.2 Essais cliniques randomisés — impossibilité générale de prévoir les effets à long terme 18

4.3 Signification statistique vs signification clinique 19

4.4 Définition de « signification clinique » 19

4.5 Difficultés relatives à l'évaluation des effets multiples 19

5. QUALITÉ DE VIE ET ÉVALUATION DES DROGUES NOUVELLES DE PRESCRIPTION 21

5.1 Définition de la qualité de vie par rapport à la santé 21

5.2 Méthodes de mesure de la qualité de vie 23

5.3 Définition de l'« utilité » 24

5.4 Mesure de l'utilité 24

TABLE DES MATIÈRES (suite)

Page

6. EXPLORATION DE LA MESURE DE LA QUALITÉ DE VIE FONDÉE SUR L'UTILITÉ DANS L'ÉVALUATION DES DROGUES NOUVELLES EN VUE DE LEUR AUTORISATION 27

6.1 Avantages des mesures de l'utilité 27

6.1.1 Possibilité de généralisation 27
6.1.2 Portée 27
6.1.3 Intégration des effets de morbidité et de mortalité 28
6.1.4 Saisie de deux dimensions de risque 28
6.1.5 Intégration des risques et avantages 28
6.1.6 Mesure des préférences pour des états de santé hypothétiques et existants 28
6.1.7 Représentation de points de vue multiples 29
6.1.8 Incorporation des préférences de temps et de risque 29
6.1.9 Bases théoriques rigoureuses 29
6.1.10 Intégration aux mesures d'évaluation économique 29
6.1.11 Intégration aux essais cliniques 29
6.1.12 Utilité pour les médecins prescripteurs 30

6.2 Inconvénients de la méthode de l'utilité dans la mesure de la qualité de vie 30

6.3 Incorporation de la mesure de la qualité de vie dans un essai de médicament : exemple 31

7. RÉSUMÉ ET CONCLUSIONS 35

8. EN CONCLUSION 37

9. BIBLIOGRAPHIE 38

Annexe A : Sélection de conférences récentes importantes sur la mesure de la qualité de vie par rapport à la santé 43

Annexe B : Principales ressources bibliographiques 44

LISTE DES FIGURES

		Page
Figure 1 :	Matrice risques-avantages	45
Figure 2 :	Boucle itérative de l'évaluation d'un médicament	46
Figure 3 :	Utilisation du thermomètre des états de santé	47
Figure 4 :	Méthode du pari type pour obtenir la valeur de l'utilité	48
Figure 5 :	Méthode du choix de temps à vivre, pour obtenir la valeur d'utilité d'un état chronique	49

LISTE DES TABLEAUX

		Page
Tableau 1 :	Quelques définitions de l' « efficacité »	50
Tableau 2 :	Taxonomie des mesures de la qualité de vie	51
Tableau 3 :	Exemple de description d'un état de santé	52
Tableau 4 :	Avantages et inconvénients des mesures de l'utilité	53
Tableau 5 :	Variations des mesures cliniques de l'état de santé, et des indices composites entre le niveau de départ et le sixième mois, par groupe de traitement (N = 303)	54
Tableau 6 :	Résultats des mesures de l'utilité et d'autres mesures	55

1. RÉSUMÉ

La grande majorité des médicaments qui offrent des avantages comportent également un certain degré de risque. Comment peut-on déterminer les médicaments qui doivent être mis à la disposition du public? De telles décisions demandent une évaluation des risques et des avantages pour la santé humaine, un processus qui est souvent appelé « analyse risques-avantages ». Le présent rapport a pour but d'étudier les méthodes actuelles et potentielles de mesure des risques et des avantages des drogues nouvelles de prescription.

Le *risque* est la combinaison de la probabilité d'un incident défavorable avec la gravité ou le danger pour la santé humaine que représente l'éventualité de cet incident. L'*innocuité* est le contraire du risque; on dit qu'un médicament est d'un emploi sûr si le risque qu'il comporte a été jugé suffisamment faible pour une situation déterminée.

L'*avantage* se rapporte à l'amélioration de la santé humaine, qui peut être attribuée au médicament. Traditionnellement, les avantages en matière de santé ont été mesurés comme des améliorations des taux de morbidité et de mortalité. La morbidité se mesure généralement par des critères de diagnostic traditionnels. Occasionnellement, on prend en considération les niveaux de fonctionnement physique, émotionnel et social. La mortalité se mesure généralement au moyen des taux de décès, en tenant compte parfois de la longévité.

L'*analyse risques-avantages* est une technique qui sert à combiner les risques et les avantages pour la santé humaine, mesurés de façon complète, et à présenter ces impacts d'une façon qui facilite la formation d'un jugement sur le bilan des risques et des avantages pour la population. La matrice risques-avantages (voir figure 1 – page 43) est une visualisation utile du processus. Ici, le risque est représenté sur un axe, tandis que l'autre axe représente l'avantage. Les médicaments à faible risque (d'un emploi sûr) seront acceptables, même avec peu d'avantages. Par contre, des médicaments à haut risque peuvent être inacceptables peu importe les avantages. Le problème est de juger les nombreux médicaments qui se situent entre les deux.

La qualité de vie en relation avec la santé représente une méthode nouvelle et globale pour mesurer l'état de santé des individus et des groupes. Ce concept peut être utilisé pour mesurer à la fois les avantages et les risques des médicaments. De cette façon, la mesure de la qualité de vie peut être utilisée comme fonction métrique d'intégration permettant de rapprocher les impacts positifs (avantages) et les impacts négatifs (risques) d'une drogue nouvelle, afin de déterminer si le bilan qui en résulte est favorable. De plus, ces mesures permettent de déterminer ce bilan du point de vue intra-patient et du point de vue inter-patients, ce qui représente une extension considérable de l'analyse risques-avantages.

Globalement, la qualité de vie est influencée par divers facteurs (revenus, satisfaction tirée du travail, perspectives sociales) qui viennent s'ajouter à l'état de santé. Dans le contexte des études cliniques, cependant, la comparaison des diverses solutions doit être orientée sur la composante de la qualité de vie qui est en relation avec la santé.

Un grand nombre de méthodes d'évaluation de la qualité de vie ont mises au point, et leur diversité est à la fois inévitable et appropriée. À l'heure actuelle, il n'existe pas de règle absolue pour mesurer la qualité de vie, et il est peu probable qu'il en existe une un jour.

Les mesures de la qualité de vie peuvent être classées en deux grands groupes : les tests spécifiques (par exemple, le questionnaire respiratoire chronique) et les tests génériques. Ces derniers se subdivisent en profils de santé (profil d'impact de la maladie) et de mesures l'utilité (test du pari type).

La méthode de l'utilité est particulièrement puissante et utile dans l'évaluation des drogues nouvelles de prescription. L'*utilité* est un concept utilisé en économie et en analyse décisionnelle, qui se rapporte au niveau de satisfaction ou de plaisir ressenti par les consommateurs. Appliquée à la santé, l'utilité se rapporte à la satisfaction que les individus tirent des différents niveaux de santé ou de perspectives de santé. Elle représente également la force des préférences des individus pour diverses perspectives finales.

Les méthodes générales qui sont employées pour obtenir des mesures cardinales de l'utilité des perspectives de santé sont maintenant bien au point. Quand elles sont adaptées et appliquées avec soin, ces méthodes de mesure sont fiables et valides. Les deux principales sont le test du « pari type » et celui de la « durée de vie équivalente »; on les fait souvent précéder de l'utilisation d'une échelle d'évaluation visuelle analogue (thermomètre des états de santé).

Les avantages de la méthode de l'utilité pour mesurer le qualité de vie sont sa possibilité de généralisation; sa globalité; sa capacité d'intégrer les effets de mortalité et de morbidité; sa capacité d'intégrer les risques et les avantages; sa capacité de faire ressortir les préférences pour les situations hypothétiques et pour les situations réelles, sa capacité de représenter des points de vue multiples; sa capacité d'utiliser divers types d'évaluateurs pour les états de santé; son incorporation des préférences de durée de vie et de risques des sujets dans l'évaluation de leur état de santé; ses bases théoriques solides; son accord avec l'évaluation économique; et sa faisabilité lorsqu'on l'utilise dans des essais cliniques. La méthode de l'utilité produit un seul score sur une échelle généralisable allant de 0,0 (mort) à 1,0 (en bonne santé). Les scores d'utilité reflètent l'information sur les choix entre les avantages d'un traitement et ses effets secondaires. Enfin, les analyses d'utilité peuvent être facilement associées à des mesures financières (en dollars) des coûts et avantages des diverses solutions de soins de santé, pour donner des analyses coûts-utilité.

La méthode de l'utilité comporte toutefois certains inconvénients. Les mesures de l'utilité ne sont pas extrêmement précises au niveau de l'individu, et demandent des mesures répétées ou des applications à des groupes afin d'atteindre un haut degré de précision. Les mesures de l'utilité exigent de longues entrevues conduites par des intervieweurs professionnels ayant reçu la formation appropriée. Les scores d'utilité ne sont pas toujours faciles à interpréter. Ces tests ne peuvent être administrés à tous les sujets. Pour finir, les scores d'utilité n'expliquent la valeur relative associée à telle ou telle perspective finale de santé; on doit alors recourir à des mesures supplémentaires pour élucider les raisons présentées.

Des mesures de l'utilité et de la qualité de vie ont été utilisées avec succès dans les essais cliniques. Par exemple, un vaste ensemble de mesures de la qualité de vie a été incorporé dans un essai multicentrique international récent sur un nouveau produit pharmacologique, l'auranofine, destiné au traitement de l'arthrite. Parmi les tests de mesure de la qualité de vie figurait la mesure de l'utilité, qui a donné des résultats particulièrement intéressants. Dans notre rapport, l'exemple de l'auranofine sera présenté en détail pour illustrer le potentiel de la méthode.

En résumé, les méthodologies permettant de mesurer la qualité de vie en relation avec la santé, y compris la méthode de l'utilité, sont maintenant suffisamment perfectionnées pour qu'il soit scientifiquement justifié de les utiliser dans des essais pharmacologiques. Les mesures de la qualité de vie ajoutent un élément d'information important et utile à la documentation traditionnellement présentée en vue de l'autorisation des médicaments. La méthode de l'utilité appliquée à la mesure de la qualité de vie est particulièrement puissante et intéressante; elle mérite donc d'être prise en considération.

2. INTRODUCTION

Étant donné que presque tous les médicaments qui ont une action favorable comportent inévitablement un certain degré de risque, comment peut-on déterminer ceux qui doivent être mis à la disposition du public? De telles décisions demandent une évaluation des risques et des avantages, un processus souvent appelé « analyse risques-avantages » (ARA). Le présent rapport étudie les méthodes actuelles et potentielles de mesure des risques et avantages des drogues nouvelles de prescription, du point de vue de la société canadienne. C'est la Direction générale de la protection de la santé (DGPS) de Santé et Bien-être social Canada qui représente les intérêts de la société en vérifiant l'efficacité et l'innocuité des médicaments.

La qualité de vie représente une perspective et un objectif importants en matière de santé. Au cours des dernières années, on a noté une très grande augmentation des activités mondiales d'élaboration et d'application de méthodes de mesure de la qualité de vie. Par exemple, on trouvera à l'annexe A une liste des dernières conférences consacrées au sujet, et à l'annexe B, les détails d'un centre d'information du gouvernement américain qui s'intéresse spécialement à ce domaine. Grâce à ces activités, un certain nombre de techniques de mesure de la qualité de vie, y compris la méthode de l'utilité, sont maintenant relativement bien au point.

Les nouvelles drogues de prescription peuvent avoir un double impact sur la qualité de vie : un impact positif, en apportant un traitement efficace; et un impact négatif, à cause de leurs effets secondaires et de réactions indésirables qu'ils entraînent. Les mesures de qualité de vie fondées sur l'utilité donnent la possibilité de mesurer les impacts positifs et négatifs avec les mêmes unités, ce qui facilite l'établissement du bilan risques-avantages. L'objet du présent rapport est d'explorer cette possibilité.

Le contenu du rapport est organisé de la façon suivante. Premièrement, un algorithme théorique, la boucle itérative de l'évaluation d'un médicament, est utilisé pour décrire les problèmes qui entourent l'évaluation des médicaments, ainsi que le mandat de la Direction générale de la DGPS et l'étendue du présent rapport. Suit une étude de l'innocuité et de l'efficacité, en termes de risques, d'avantages et d'analyse risques-avantages, puis des techniques de mesure de la qualité de vie récemment mises au point. Pour finir, est étudiée l'application de la méthode de l'utilité par la mesure de la qualité de vie en relation avec la santé, comme méthode pouvant aider à établir le bilan risques — avantages, cette méthode est illustrée par l'exemple de l'essai clinique d'un sel d'or oral (auranofine) destiné aux patients arthritiques.

3. ANALYSE RISQUES-AVANTAGES DES DROGUES NOUVELLES

3.1 Boucle itérative d'évaluation d'un médicament

Le point de vue adopté dans le présent rapport sera celui de la société canadienne. Au Canada, c'est la Direction générale de la protection de la santé (DGPS), de Santé et Bien-être social Canada1, qui a la responsabilité2 d'assurer l'innocuité et l'efficacité des drogues nouvelles. On peut dire que le rôle de la DGPS est d'évaluer l'innocuité et l'efficacité théorique et pratique des drogues nouvelles, du point de vue de la société, en employant une version adaptée de la boucle itérative de mesure (Tugwell et al, 1985) et la boucle itérative d'évaluation de technologie (Tugwell et al, 1986) représentée à la figure 2 (page 44). Cette boucle est un modèle simple et schématique qui peut servir à rassembler et à classer l'information. En plus de décrire le mandat de la DGPS, la boucle itérative d'évaluation d'un médicament servira à délimiter l'étendue de notre rapport.

La boucle itérative d'évaluation d'un médicament offre une méthode systématique pour rassembler les sous-ensembles spécifiques d'information en matière de santé qui sont nécessaires pour prendre une décision rationnelle sur l'adoption et l'utilisation des drogues nouvelles. Ce système subdivise le spectre d'information en sous-groupes qui constituent une progression logique : on part de la quantification du fardeau de la maladie pour en arriver à vérifier l'innocuité d'emploi du produit et sa capacité de prévenir ou de réduire la maladie; puis vient la diffusion du produit, et, en fin de boucle, on détermine si le fardeau de la maladie a été réduit.

Chacune des étapes de la boucle d'évaluation d'un médicament soulève différentes questions ayant trait à la recherche ou à l'évaluation. Par exemple, le système distingue les concepts suivants : l'innocuité d'un médicament, son efficacité théorique, son efficacité pratique et sa rentabilité. La conception de la boucle fait ressortir l'importance de la surveillance à exercer après que le produit a été approuvé et mis sur le marché (pharmacovigilance). Le processus de réduction du fardeau de la maladie est également itératif. Dans la plupart des situations de soins de santé, le fardeau de la maladie n'est réduit que par étapes, et des ajustements répétés sont nécessaires pour soulager même la portion pour laquelle existent des interventions favorables (Tugwell et al, 1985, 1986).

---

[1] Bien que ce rapport traite spécifiquement du mandat du Bureau des médicaments humains prescrits de la Direction des médicaments, les sujets abordés et les recommandations s'appliquent, en grande partie, à d'autres bureaux et directions.

[2] *Loi sur les stupéfiants*, *Loi sur les aliments et drogues* et leurs règlements d'application.

Le mandat de la DGPS est axé sur l'étape 3 de la boucle (innocuité et efficacité), mais englobe aussi l'étape 2 (exactitude du dépistage et du diagnostic) et l'étape 4 (efficacité dans la communauté). Dans une certaine mesure, la DGPS est également intéressée à l'étape 7 (surveillance et réévaluation). Mais elle n'a pas pour mandat d'envisager la rentabilité d'un produit (c'est-à-dire son coût comparé à celui d'autres produits), bien que la question puisse être d'un certain intérêt pour des pharmacopées particulières (au niveau des hôpitaux ou des provinces), les patients ou les médecins individuels, une fois que le produit a été approuvé pour la mise sur le marché par la DGPS.

Voici la description des sept étapes de la boucle itérative d'évaluation d'un médicament.

- **Étape 1 : Fardeau de la maladie**

  Quand on évalue une drogue nouvelle du point de vue de la société, la première étape est de déterminer le fardeau de la maladie que le médicament a pour but de soulager. Le fardeau de la maladie peut être évalué en déterminant les niveaux actuels de morbidité (c'est-à-dire fonctionnement physique, fonctionnement émotionnel, fonctionnement social, symptômes et jours d'incapacité) ou de mortalité (c'est-à-dire taux de décès et années de vie perdues). Les instruments de mesure utilisés pour déterminer le fardeau de la maladie doivent être adaptés à cette maladie, précis et faciles à utiliser et à interpréter. En plus d'évaluer le fardeau de la maladie, il est souhaitable de comprendre ce qui cause la maladie, puis de déterminer si la maladie est modifiable par un traitement médicamenteux. Les faits pouvant établir les causes des problèmes de santé contribuant au fardeau de la maladie peuvent être évalués en suivant les lignes directrices comme celles qui ont été élaborées par le Service d'épidémiologie clinique et de biostatistique de l'Université McMaster (1981).

- **Étape 2 : Dépistage et exactitude du diagnostic**

  La deuxième étape vise à établir les preuves de la spécificité, de la sensibilité et de l'utilisation des tests et des médicaments spécifiques pour établir le diagnostic de la maladie. Ceci comprend le problème de la spécificité et de la sensibilité des programmes de dépistage et des autres études conduites par les professionnels des soins de santé pour déterminer les personnes qui pourraient bénéficier d'un traitement médicamenteux et celles qui ne pourraient pas en bénéficier.

- **Étape 3 : Innocuité et efficacité théorique**

  Cette étape a pour but de déterminer si le médicament a une action favorable. Fait-il plus de bien que de mal aux patients, dans des conditions idéales (c'est-à-dire exactitude optimale du diagnostic et conformité du personnel soignant et des patients aux recommandations applicables)?

- **Étape 4 : Efficacité pratique, dans la communauté**

  L'efficacité pratique a rapport à l'utilité. Par exemple, dans des conditions de vie réelles, le médicament fait-il plus de bien que de mal aux gens auxquels il est offert? La réponse ne dépend pas seulement de l'exactitude du diagnostic (étape 2 ci-haut), ni de l'efficacité du produit (étape 3 ci-haut), mais également de la conformité au traitement du personnel soignant et des patients, ainsi que du degré de couverture du médicament. La conformité au traitement du personnel soignant est liée à un diagnostic approprié et à une bonne conduite du traitement (conformité aux recommandations d'utilisation du produit). La fidélité des patients au traitement dépend de la façon dont ils se conforment aux recommandations et au traitement du personnel soignant. La facilité de prise d'un produit, les effets secondaires qui lui sont attribuables, le coût du produit pour le patient, ou la mesure dans laquelle le patient peut comprendre les avantages potentiels de la prise du médicament sont tous des facteurs qui jouent un rôle dans la conformité du patient au traitement.

  Le degré de couverture se rapporte à la mesure dans laquelle le produit est utilisé de façon appropriée par tous ceux qui pourraient en bénéficier. L'utilisation dépend de la disponibilité (le produit est-il facile à obtenir?) et de son degré d'acceptation (ressources, installations, comportement des professionnels de la santé et avantages escomptés du service de santé). Pour assurer un degré de couverture acceptable, le programme de traitement médicamenteux doit être en rapport avec un besoin chez une population définie, pendant une période déterminée (Tugwell, 1985).

- **Étape 5 : Rentabilité**

  Cette étape consiste à déterminer s'il est possible d'administrer un traitement médicamenteux efficace d'une manière telle que les avantages du programme (c'est-à-dire le nombre de vies ou d'années de vie sauvées, le nombre de journées d'incapacité évitées, l'amélioration de la qualité de vie, la réduction des coûts de traitement futurs) l'emportent sur les coûts (coût du dépistage, du traitement et des effets secondaires, et impact des effets secondaires sur la qualité de vie).

  Les méthodes coûts-avantages, coûts-efficacité et coûts-utilité qui servent à mesurer la rentabilité sont étudiées dans Drummond et al, (1987).

À la différence de l'innocuité, de l'efficacité théorique et de l'efficacité pratique d'un traitement médicamenteux, il est clair que l'évaluation de sa rentabilité ne fait pas partie du mandat de la DGPS. Pour mesurer cette rentabilité, nous comparons les coûts et avantages différentiels du produit à ceux d'un autre traitement du même état, et à ceux d'une autre utilisation des ressources. La rentabilité d'un traitement médicamenteux n'est pas statique : elle doit être réévaluée en fonction des nouveaux produits, des nouvelles technologies et des modifications de l'environnement. La rentabilité est un problème qui intéresse d'une façon générale la société, et plus spécialement ceux qui peuvent avoir à payer le nouveau traitement, c'est-à-dire les patients individuels, les hôpitaux, les pharmacopées provinciales et privées. Les considérations relatives à la rentabilité ne font pas partie de la présente étude.

- **Étape 6 : Synthèse et mise en application**

  La sixième étape intègre la faisabilité avec les estimations de l'efficacité et de la rentabilité dans la communauté obtenues dans les étapes précédentes, pour faire les recommandations de mise en application d'une intervention. Cette étape permet de déterminer les contraintes pouvant limiter les estimations d'efficacité pratique et de rentabilité qui existent dans le cadre où l'intervention sera mise en application. Elle consiste également à déterminer si, et dans quelle mesure, ces contraintes peuvent être éliminées ou réduites, pour aboutir à une conclusion sur l'impact probable du médicament sur le fardeau de la maladie. Les contraintes comprennent les barrières sociales, culturelles et politiques, ainsi que le caractère adéquat et la disponibilité des installations et de la main-d'oeuvre, et, pour finir, les considérations budgétaires. Il est indispensable dans cette étape d'établir des objectifs et des critères d'évaluation réalistes (Tugwell, 1985).

- **Étape 7 : Surveillance et réévaluation**

  Cette étape consiste à évaluer comment le nouveau traitement médicamenteux a permis, à court, à moyen et à long terme de réduire le fardeau de la maladie. La sélection d'indicateurs servant à surveiller le succès du nouveau produit nécessite la prise en considération du spectre de la maladie, des groupes d'âge, de l'exactitude et de la faisabilité (Tugwell, 1985).

  La boucle s'achève en retournant au fardeau de la maladie, que l'on espère réduit. Chaque circuit de la boucle donne des informations sur la façon dont le produit peut continuer à réduire le fardeau de la maladie.

3.1.1 Portée du rapport

La DGPS s'intéresse directement à un certain nombre d'étapes de la boucle : l'étape 2 (exactitude du diagnostic), l'étape 3 (innocuité et efficacité), l'étape 4 (efficacité dans la communauté) et l'étape 7 (surveillance et réévaluation). La DGPS s'occupe en effet de l'évaluation de l'innocuité des substances utilisées pour le diagnostic des maladies, comme les opacifiants radiologiques. Une partie importante des observations soumises à la DGPS par les compagnies pharmaceutiques vient des essais en phase III conduits dans des cadres réalistes, fournissant ainsi des observations destinées à l'étape 4 (efficacité pratique dans la communauté). La surveillance après commercialisation, ou pharmacovigilance, comme le *Federal Voluntary Drug Adverse Reaction Reporting Program*, fait partie de l'étape 7. La décision d'autorisation, cependant, repose largement sur les tests d'efficacité, les observations soumises à l'étape 3, et l'objectif essentiel de la DGPS (innocuité et efficacité). Le reste de la présente étude portera sur le rôle de la DGPS aux étapes 3 et 4 dans l'évaluation du rapport risques-avantages des nouveaux traitements médicamenteux.

3.2 Risques, avantages et analyses risques-avantages : définitions

3.2.1 Risque (innocuité)

Les organisations *publiques* chargées de la réglementation des risques emploient le terme *innocuité* de préférence au terme *risques*, qui est négatif. Il n'en reste pas moins que risque et innocuité sont les deux faces du même problème et qu'ils correspondent à des concepts relatifs : aucun produit n'est ni tout à fait sûr ni sans risque. Intuitivement, un médicament sûr est un médicament qui ne comporte pas de dangers. Ici, les termes « sûr » et « innocuité » correspondent à un degré de risque qui a été jugé si faible qu'il est négligable (O'Brien, 1986).

Le *Petit Robert* (1988) défini le risque comme suit :

> « ... danger éventuel plus ou moins prévisible ... »

Le risque peut être considéré comme une notion associant la « probabilité » ou les chances de survenue d'un événement, et la « gravité » ou le danger que présenterait cet événement s'il se produisait (Lowrance, 1976, p. 94). L'annexe B fournit une bibliographie complète des ouvrages sur le risque. La motion de risque suppose que les investigateurs et les décideurs doivent se préoccuper à la fois de la nature du risque et de la probabilité de sa survenue. La *Royal Society of Great Britain* (1983) donne une définition du risque qui correspond à ce point de vue. Selon elle, le risque est :

> « ... la probabilité qu'un incident "défavorable" particulier se produise pendant une période de temps déterminée ou résulte d'une contrainte particulière ».

L' adjectif défavorable dans cette définition fait ressortir que l'événement est indésirable, qu'il a une connotation de perte possible, de danger, de blessure ou de dommage.

L'*Office of Technology Assessment* (1978) donne une définition du risque qui est plus explicite :

> « ... la probabilité que se produise un événement défavorable ou indésirable et la gravité du dommage qui en résulterait pour la santé des individus d'une population définie, associées à l'utilisation d'une technologie médicale appliquée à un problème médical donné dans des conditions d'utilisation déterminées. »

Cette définition peut aussi être appliquée à l'évaluation des médicaments. Elle a une grande portée, car elle comprend plusieurs facteurs : le problème médical auquel s'applique la technologie ou le médicament en cours d'évaluation, la population touchée, et les conditions d'utilisation dans lesquelles la technologie (le médicament) est appliquée. La définition couvre le risque dans un cadre idéal (recherche) et dans des cadres de niveau moyen et de type courant (communauté). Normalement, lorsqu'on parle de jugements « d'efficacité et d'innocuité », on suppose que le risque est mesuré dans des conditions d'utilisation idéales.

Compte tenu de la définition du risque donnée ci-dessus, l'innocuité peut être définie comme suit :

> « ... un jugement de l'acceptabilité du risque dans une situation déterminée », (Banta 1981, p. 99).

3.2.2 Avantages (efficacité théorique/efficacité pratique/rentabilité)

La DGPS demande plus que la démonstration de l'innocuité d'un produit. Avant que la DGPS n'autorise la mise en marché du produit, on doit également prouver qu'il est efficace, soit directement, soit indirectement. Comme l'innocuité, l'efficacité est un concept relatif pour lequel les définitions ne manquent pas (tableau 1, page 48). Selon Banta (1981), il existe quatre facteurs qui, pris ensemble, donnent une vision complète du concept. Ces facteurs sont les suivants :

a) l'avantage à obtenir,

b) le problème médical donnant lieu à l'utilisation de la technologie (médicament, etc.),

c) la population atteinte, et

d) les conditions d'utilisation dans lesquelles la technologie (ou le médicament) est appliquée.

Le mot « avantage » est en corrélation avec les termes suivants : « efficacité théorique » (avantages dans des conditions d'utilisation idéales), « efficacité pratique » (avantages dans des conditions d'utilisation moyennes) et « rentabilité » (avantages comparés au coût).

Les résultats servant de critères pour mesurer les avantages ont généralement été limités à la morbidité et à la mortalité; on a donné moins d'importance à l'espérance de vie (longévité) ou aux facteurs psychosociaux et fonctionnels. La définition de l'avantage dépendra des objectifs de l'investigateur et du type de produit évalué. Pour une analyse récente des différentes façons de mesurer les avantages cliniques, voir Laupacis, Sackett et Roberts (1988). Ces auteurs examinent les avantages et les inconvénients de mesures comme la réduction du risque relatif et du risque absolu, et le nombre nécessaire de personnes à traiter du point de vue d'un clinicien évaluant le risque individuel.

3.2.3 Analyse risques-avantages

En 1971, Feinstein indiquait qu'il n'existe pas de procédure standard satisfaisante pour évaluer l'innocuité et l'efficacité des médicaments. Actuellement, la terminologie de l'analyse risques-avantages, plus généralement connue sous le nom d' « évaluation du risque », est encore loin d'être normalisée (Jones et Akehurst, 1980).

Le Conseil du Trésor du Canada (1979, p. 10) définit l'analyse risques-avantages comme :

> « ... un cas spécial (d'analyse coûts-avantages) où une attention explicite est portée aux catégories de coûts humains (risque de perte de vie ou de membre) pour lesquels n'existent pas encore de techniques d'évaluation établies. Dans l'analyse risques-avantages, les avantages peuvent être interprétés dans un sens net (le bilan des avantages ordinaires et des coûts ordinaires), mais les risques de perte de vie et de souffrance humaine représentent des coûts extraordinaires qui sont à évaluer séparément. »

Dans cette définition, le terme « coûts » ne se rapporte pas à des coûts en argent, mais à des coûts « humains », c'est-à-dire à des augmentations de probabilité de détérioration de l'état de santé. Par conséquent, l'analyse risques-avantages est une méthode qui permet d'analyser de façon globale tous les impacts sur la santé humaine, qu'ils soient positifs ou négatifs, et de présenter ces impacts d'une façon qui facilite l'obtention d'un jugement sur le bilan des risques et avantages pour la population.

La matrice risques-avantages de Teeling-Smith, à la figure 1 (page 43), donne une représentation visuelle utile du problème. Dans cette matrice, l'innocuité est représentée sur l'axe vertical, et la signification clinique et sociale sur l'axe horizontal. Un médicament qui se trouve dans la partie inférieure de la matrice est acceptable parce qu'il est sûr, tandis qu'un médicament situé dans le coin supérieur gauche sera inacceptable, parce qu'il donne un faible avantage au prix d'un risque élevé. La difficulté consiste à déterminer la façon de placer les médicaments sur cette matrice. Les produits dont la position sur la matrice dépend de la façon dont ils sont utilisés pose un problème particulier. Par exemple, le chloramphénicol se trouve dans le coin supérieur droit lorsqu'il est utilisé contre la typhoïde, mais se déplace vers le coin supérieur gauche quand il est prescrit pour une irritation de la gorge. Un produit comme le bénoxaprofène, dont l'innocuité et l'efficacité sont importantes pour de jeunes arthritiques invalides, peut perdre ses deux caractéristiques s'il est administré à des arthritiques âgés (Teeling-Smith, 1983).

3.3 Méthodes de mesure des risques-avantages

Jones et Akehurst (1980) étudient la méthode d'évaluation de risque et la décomposent en plusieurs stades : identification, quantification et évaluation du risque. Étudions ces divers stades dans l'analyse risques-avantages.

3.3.1 Identification

Le premier stade, l'« identification » de l'ensemble des événements à risque associés à un certain ensemble de circonstances, est relativement peu compliqué. Il s'agit tout simplement de dresser une liste de tous les événements à risque possibles. Cependant, les détails de la liste ainsi compilée dépendront de la perception des risques de celui qui le compile. De plus, pour certaines drogues nouvelles, les risques peuvent ne pas avoir encore été identifiés.

3.3.2 Quantification

La « quantification du risque », deuxième stade de la méthode, présente deux difficultés. La première, c'est que les événements à risque peuvent être complexes et demander l'estimation de plusieurs probabilités de risque en interrelation. Des effets multiples peuvent entraîner des disparités ou des contre-indications. La deuxième difficulté, c'est qu'il puisse y avoir trop peu de données convenables pour qu'on puisse en déduire une estimation des probabilités.

La méthode habituellement employée pour quantifier les risques consiste à mettre en tableau et à comparer le nombre d'effets secondaires ou de réactions indésirables qui se produisent chez les patients traités avec le médicament d'essai ou un placebo. Laupacis, Sackett et Roberts (1988) font remarquer qu'en l'absence d'évaluation de qualité de la vie qui révèle à la fois les avantages et les inconvénients des traitements chez tous les patients, la nocivité absolue d'un traitement peut être quantifiée de la même façon que l'avantage absolu : en soustrayant la nocivité dans le groupe témoin de celle dans le groupe des sujets traités. Mais ces méthodes ne permettent pas de résoudre la question de savoir dans quelle mesure une réaction indésirable est une cause de préoccupation. Jusqu'où doivent aller les effets bénéfiques pour que l'efficacité soit établie? Dans le cas des réactions indésirables, on se sert des suppositions prudentes du « pire cas », mais dans quelle mesure une supposition prudente l'est-elle vraiment? Ici, les suppositions ne sont pas explicites. On ne dispose pas souvent de critères opérationnels pour identifier les réactions indésirables (Feinstein, 1971).

Une autre méthode pour traiter le problème des effets multiples consiste à se servir d'une analyse par arbre de décision. L'« arbre » permet de décomposer des dangers complexes en une série de possibilités plus simples pour lesquelles on dispose en principe d'estimations de probabilité de survenue. Cette méthode est décrite par Fineberg (1980) et Weinstein et Fineberg (1980).

Le deuxième problème — un manque de données convenables sur lesquelles fonder les estimations de probabilités de risque — est plus grave lorsqu'on considère des événements rares. Dans la plupart des études de quantification du risque, l'intégration subjective humaine, généralement des jugements d'« experts », sera incorporée dans une importante mesure (Jones et Akehurst, 1980). Des techniques rigoureuses, comme l'« encodage de probabilité » (Spetzler et von Holstein, 1975), existent pour quantifier les jugements de probabilité des experts dans de telles circonstances. On peut également utiliser les méthodes bayésiennes pour les études d'estimation du risque (Altshuler, 1976). Finalement, des techniques comme la méthode Delphi (Linstone et Twoff, 1975) peuvent être utilisées pour concilier les divergences d'opinions.

3.3.3 Évaluation

Le troisième stade de la détermination du risque est l'évaluation du risque ou l'analyse risques-avantages. Explicitement ou implicitement, au moment de prendre des décisions sur leur comportement futur, les individus procèdent à une estimation des risques par rapport aux avantages possibles. De telles appréciations sont également effectuées par la société, par l'intermédiaire d'agents ayant la confiance du public, comme la DGPS.

S'il existe une documentation considérable sur l'analyse risques-avantages, il y a peu d'exemples de cas où des méthodes analytiques ont servi de base aux décisions d'autoriser un produit. Un examen de la documentation indique la nécessité de séparer les jugements scientifiques des jugements de valeur.

Des indices thérapeutiques ont été utilisés pour décrire scientifiquement la relation entre les doses d'un produit qui sont nécessaires pour produire des effets indésirables et celles qui sont nécessaires pour produire des effets désirés. L'index thérapeutique se définit généralement comme le rapport entre la dose toxique médiane et la dose efficace médiane (TD50/ED50), ou entre la dose létale médiane et la dose efficace médiane (LD50/ED50). Goodman et Gilman (1985) ont noté qu'un produit n'a pas un seul index thérapeutique, mais plusieurs. La marge de sécurité de l'aspirine dans son action contre les maux de tête est plus grande que sa marge de sécurité dans son action contre les douleurs arthritiques. De la même façon, on peut calculer plusieurs indices thérapeutiques correspondant à chacun des effets désirés. Un opiacé de synthèse peut provoquer moins de constipation que la morphine et ne présenter cependant aucun avantage sur la molécule mère en ce qui concerne la dépression respiratoire ou la sédation. De plus, un produit peut être sélectif dans un contexte, et ne pas l'être dans un autre. Par exemple, les antihistaminiques sont des antagonistes sélectifs de l'histamine, mais deux seulement de ces produits ont cet effet périphérique sans provoquer de sédation centrale significative. Finalement, un produit peut avoir une marge de sécurité adéquate chez la plupart des patients, mais cela ne signifie rien pour celui qui présente une réaction inhabituelle au produit. Par exemple, la pénicilline est essentiellement non toxique chez la grande majorité des patients, et pourtant elle peut provoquer la mort chez ceux qui y sont devenus allergiques. L'index thérapeutique donne donc une information très limitée aux fins d'autorisation.

Banta (1981) décrit cinq techniques permettant d'évaluer l'innocuité et l'efficacité.

- Les techniques **précliniques** sont des tests biochimiques ou animaux qui servent de guide pour définir l'activité thérapeutique et la toxicité potentielles d'un médicament. La certitude avec laquelle les modèles animaux permettent de déterminer les effets probables des produits sur l'homme est un problème controversé. D'autres auteurs ont traité la question d'une façon plus détaillée (Laurence, McLean et Weatherall, 1984; Zbinden, 1982).

- Les techniques **non rigoureuses**, selon Benta, comprennent l'expérience personnelle, dominée par des impressions qualitatives; l'expérience des pairs, qui se dégage par l'échange de communications personnelles, d'articles de revues et par consensus professionnel, soit rigoureux, soit non rigoureux (Banta, 1981, p. 104).

- L'**épidémiologie** est l'étude des déterminants et de la distribution des maladies et des traumatismes dans les populations. Les méthodes épidémiologiques et statistiques sont fréquemment utilisées dans les modèles informatiques et les techniques de régressions multiples pour étudier l'efficacité des interventions médicales.

- Dans les **essais cliniques randomisés**, les sujets sont répartis de façon aléatoire dans un groupe expérimental et dans groupe témoin. Le groupe témoin reçoit un placebo ou un traitement standard établi. Le groupe expérimental reçoit le nouveau traitement. Les désavantages des essais cliniques randomisés sont étudiés au chapitre 4. Parmi les avantages, on peut citer l'élimination du biais lié à l'assignation du traitement, la prévention du biais causé par les variables qui existent dans l'expérimentation, mais qui ne sont pas prises en considération dans la conception, et la validité des tests de signification statistiques qui sont utilisés pour comparer le traitement (Banta 1981, p. 107).

- La **synthèse rigoureuse** est utilisée quand les données provenant des techniques ci-dessus ne conduisent pas à une décision sans équivoque. Cette synthèse comporte une compilation de toutes les informations disponibles. Un modèle approprié de cette synthèse est la boucle itérative d'évaluation d'un médicament examinée précédemment.

3.4 Autorisation des nouveaux produits de prescription

Le présent chapitre donne un bref aperçu des études entreprises et des données recueillies dans le cadre du mécanisme d'autorisation et de surveillance des drogues nouvelles.

3.4.1 Revue des conditions requises pour l'autorisation des médicaments au Canada

Un essai clinique de drogue nouvelle compte quatre phases.

- **Phase I** — Une consiste à effectuer l'essai initial d'un médicament expérimental chez des sujets volontaires normaux ou des patients en assez bonne santé afin de déterminer l'innocuité et les effets pharmacologiques (p. ex., des réactions indésirables cliniquement importantes, des signes physiques et/ou des anomalies de laboratoire), la fréquence de la posologie et la durée de l'exposition à un médicament. Des études portant sur un nombre réduit de patients peuvent être conduites pour évaluer l'efficacité clinique d'un médicament. D'une façon générale, le nombre total de sujets auxquels le médicament est administré est d'environ 80.

- **Phase II** — Ce sont des études initiales conçues pour évaluer l'efficacité d'un produit chez des populations sélectionnées de patients à qui le médicament est destiné. Les informations concernant l'innocuité relative sont également réunies dans les études de phase II. Les études visent à déterminer, par exemple, le moment où un produit provoque une modification spécifique escomptée pour un paramètre de laboratoire. D'une façon générale, le nombre total de patients à qui le médicament a été administré est d'environ 200.

- **Phase III** — Il s'agit d'essais cliniques contrôlés et non contrôlés conduits chez les populations de patients auxquelles le produit est finalement destiné. Ces études fournissent des informations sur l'innocuité et l'efficacité du produit chez un nombre relativement important de patients dans des conditions normales d'utilisation. Elles fournissent une bonne partie de l'information qui est nécessaire pour le prospectus d'emballage du produit et son étiquetage. En général, cette phase comporte l'administration du produit à un nombre de patients variant entre 500 et 3 000.

- **Phase IV** — Des études d'évaluation après commercialisation, ou de pharmacovigilance, sont souvent conduites afin d'obtenir des détails supplémentaires pour une meilleure information sur l'efficacité et/ou le profil d'innocuité du produit. On peut alors évaluer différentes formulations, posologies et durées de traitement, ainsi que les interactions médicamenteuses. On peut également étudier de nouveaux groupes classés par âge, sexe, race ou autres types de patients. Ces études peuvent apporter des renseignements sur des réactions indésirables rares ou peu fréquentes, ainsi que révéler des effets bénéfiques jusque là inconnus. Au Canada, des études de pharmacovigilance ne sont pas obligatoires pour l'obtention ou le maintien d'une autorisation de commercialiser un produit. Cependant, la compagnie qui commercialise le produit doit signaler à la DGPS toute réaction indésirable se produisant au Canada. En pratique, tous les renseignements obtenus à l'échelle mondiale sur les réactions indésirables sont communiqués à la DGPS.

3.4.2 Approches actuelles utilisées par les compagnies pharmaceutiques

Ce sont les essais de phase III qui apportent une grande partie des renseignements importants pour les analyses risques-avantages; ils formeront donc le sujet de la présente section. Cependant, la plus grande partie de la discussion s'appliquera aux autres phases de la mise au point du produit.

Dans le choix des critères d'innocuité et d'efficacité à appliquer dans les essais, on utilise généralement les mesures d'efficacité qui sont bien établies pour la ou les maladies à l'étude. Les tests doivent également être évalués en fonction des ressources disponibles, et être efficaces et rentables par rapport au but visé. Les utilisations prévues d'un produit (usage externe ou parentéral, pour des patients généralement sains ou gravement malades, capable de sauver des vies ou avec un effet thérapeutique minimal), l'expérience antérieure tirée de l'utilisation du produit (profil toxicologique) ou d'autres produits similaires, tels sont les éléments qui détermineront les tests d'innocuité spécifiques à incorporer dans une étude. Il n'existe pas de séries normalisées de paramètres de laboratoire qui soient évaluées dans toutes les études cliniques, pas plus qu'il n'existe une norme qui s'applique à tous les produits. Mais il existe des normes flexibles et des lignes de conduite générales en ce qui concerne les paramètres de base dans les domaines de l'hématologie, de la chimie clinique, des analyses d'urine et de l'innocuité oculaire.

Divers types de critères peuvent être utilisés pour mesurer les paramètres d'efficacité (ou d'innocuité) :

a) les critères de présence ou d'absence (symptômes, signes, lésions ou autres manifestations de la maladie qui sont présents ou non);

b) critères gradués ou échelonnés (comme les échelles visuelles analogues);

c) critères de variation relative (mesures de tests spécifiques donnant des indications directes ou indirectes d'efficacité); ou

d) critères globaux (évaluation globale de la maladie d'un patient, ou des variations de sa gravité, fondée sur des critères spécifiques ou généraux, ou un système de notation destiné à évaluer chacun des facteurs) (Spilker, 1984).

4. LIMITES DU SYSTÈME ACTUEL D'ÉVALUATION DU RAPPORT RISQUES-AVANTAGES DES DROGUES NOUVELLES

Le système actuel fournit une structure permettant d'assurer l'innocuité et l'efficacité des drogues nouvelles avant d'en autoriser la commercialisation à grande échelle au Canada. Le processus d'autorisation des médicaments incorpore cependant à la fois les avantages et les limites des méthodologies disponibles. Le présent chapitre met en évidence certaines des limites de ces méthodologies.

4.1 Essais cliniques randomisés — limites dues à la taille des échantillons

Les essais cliniques randomisés prévoient des tailles d'échantillons destinées à établir l'efficacité théorique ou pratique d'un produit pour l'ensemble des patients. Ces tailles d'échantillons sont généralement trop réduites pour qu'on puisse prévoir les événements rares.[3] De plus, les résultats d'un essai ne peuvent être généralisés pour des populations de patients autres que celles qui sont incluses dans l'essai.

4.2 Essais cliniques randomisés — impossibilité générale de prévoir les effets à long terme

Dans la plupart des essais cliniques randomisés, le point final correspond à des effets à court ou à moyen terme. Un suivi à long terme étant rarement inclus dans un essai clinique randomisé il est difficile de prévoir les effets à long terme d'un médicament. L'importance d'un suivi à long terme dans l'évaluation de l'innocuité d'un produit est illustrée de façon décisive par l'essai de l'OMS sur la prévention primaire des troubles de l'ischémie cardiaque. Dans cet essai, un groupe d'hommes en bonne santé, au taux de cholestérol sérique modérément augmenté et traités avec du clofibrate pour diminuer ce cholestérol, avait présenté une diminution significative du risque d'infarctus du myocarde non mortel,

---

* Si, par exemple, un produit cause une réaction indésirable grave chez 1/1000 (0,1 % ou 0,0001) des patients qui le prennent, une « règle de trois » nous dira que pour trouver au moins une réaction indésirable avec un degré de confiance de 95 %, nous devrons suivre (1/00, 001) X 3 ou 3 000 patients, ce qui représente un nombre beaucoup plus important que n'en comporte pratiquement tout essai clinique. Bien que les essais randomisés représentent la seule manière fiable de déterminer l'efficacité d'un produit, une surveillance après commercialisation (pharmacovigilance) est nécessaire pour en déterminer l'innocuité (Sackette, Haynes et al, 1980).

par comparaison aux témoins placebo. Mais un suivi (Comité des investigateurs principaux, 1980) a démontré par la suite une augmentation persistante et significative de la mortalité globale dans le groupe de patients traité au clofibrate. Les essais qui ne portent pas sur des échantillons de taille suffisante, ou qui ne sont pas suivis assez longtemps pour déterminer les effets de mortalité globale, peuvent entraîner des conclusions faussement négatives.

4.3 Signification statistique vs signification clinique

Le choix d'un niveau de signification statistique $p = 0,05$ est arbitraire. En effet, une différence qui est « statistiquement significative » peut être sans importance clinique. La question « dans quelle mesure une action est-elle pratiquement efficace? » a besoin d'être définie avant le commencement de l'étude. Par ailleurs, l'importance clinique est aussi un concept relatif, et les cliniciens sont souvent en désaccord sur le degré de différence qui est important. Les risques relatifs, les rapports de probabilité, les réductions du risque absolu et le nombre de patients à traiter sont des méthodes cliniques qui visent à résoudre ce problème (Laupacis, Sackett et Roberts, 1988).

4.4 Définition de « signification clinique »

Les résultats peuvent varier en fonction de la mesure du résultat choisi. Une récente étude consacrée à la comparaison de soins médicaux seuls vs des soins chirurgicaux, pour des patients à haut risque d'accident vasculaire cérébral, n'a pas permis de déterminer d'augmentation ou de diminution significative du risque d'accident vasculaire cérébral. Cependant, les données d'états fonctionnels recueillies pendant l'essai ont montré que la chirurgie entraînait une détérioration transitoire significative de l'état fonctionnel par comparaison aux malades traités médicalement (Haynes, Mukherjee, Sackett et al, 1987). Cet exemple montre bien que la définition des points finaux d'importance clinique (dans ce cas, l'accident vasculaire cérébral vs la capacité fonctionnelle) définit de façon critique si les résultats sont ou non significatifs.

4.5 Difficultés relatives à l'évaluation des effets multiples

Il est difficile d'évaluer des effets multiples et souvent contradictoires. Par exemple, les points finaux qui servent à déterminer l'efficacité pharmacologique vs l'efficacité thérapeutique sont différents, ce qui les rend difficiles à intégrer dans les termes d'un rapport risques-avantages global. La même observation s'applique également aux effets immédiats vs ceux à court terme et vs ceux à long terme. Des symptômes comme la douleur, l'inconfort ou l'incapacité sont souvent délibérément retirés de l'évaluation, parce

que ces données sont considérées comme trop imprécises pour une attention scientifique, même si elles sont plus appropriées pour des objectifs cliniques. Par exemple, si l'on s'entend généralement pour admettre que l'impact du traitement sur la famille du patient n'est pas sans importance, il n'en reste pas moins qu'on le mesure rarement. Des facteurs comme un régime alimentaire plus agréable et un changement de fréquence des injections ont également une nette influence sur l'observance du traitement par le patient et, par conséquent, sur l'efficacité pratique du produit dans la communauté.

Une méthode couramment employée pour évaluer le rapport risques-avantages consiste à faire la somme du nombre d'avantages et celle du nombre des effets secondaires (Fineberg, 1971). Le problème est de déterminer :

a) les effets à inclure,

b) la façon de les définir et de les quantifier, et

c) combien d'effets doivent être quantifiés et dans quelle proportion.

Les réponses varieront en fonction de la nature du traitement. Compte tenu de l'absence presque totale de mesures normalisées, il est difficile d'établir un niveau de risque acceptable.

5. QUALITÉ DE VIE ET ÉVALUATION DES DROGUES NOUVELLES DE PRESCRIPTION

5.1 Définition de la qualité de vie par rapport à la santé

La qualité de vie est un concept large et multidimensionnel. Les termes employés varient du niveau familier à celui des termes techniques, avec des significations nettement précisées. Si l'expression « qualité de vie » se rapporte pratiquement toujours à la satisfaction et au plaisir des êtres humains dans une situation ou un environnement particulier, la signification spécifique de la qualité de vie dans diverses applications et disciplines varie considérablement.

On peut évaluer la qualité de vie au moyen du revenu réel par habitant — ce qui correspond à une évaluation des possibilités de consommation ou du niveau de vie matériel. Dans les ouvrages sur les indicateurs sociaux, la notion de qualité de vie recouvre un territoire plus vaste dans lequel les dimensions d'ordre économique sont complétées par des renseignements sur les équipements recréatifs et culturels locaux, les possibilités de formation scolaire, la qualité de l'eau et de l'air, l'importance de la criminalité, ainsi que la nature et le niveau des services et des activités communautaires.

Ces mesures sont valables et utiles pour les fins auxquelles elles ont été créées. Dans le contexte de l'évaluation des soins de santé, toutefois, ces notions de qualité de vie sont trop vastes. Les mesures qui visent plus particulièrement les aspects de la qualité de vie susceptibles d'être influencés par les interventions dans le domaine des soins de santé sont appelées mesures de la qualité de vie liées à la santé.

Les interventions en matière de soins de santé ont eu principalement pour objet, par le passé, de réduire la mortalité. Toutefois, la société se préoccupe de plus en plus non pas seulement de l'accroissement de la longévité, mais aussi de l'amélioration de la qualité de vie. La qualité de vie liée à la santé est apparue comme une étiquette pouvant être appliquée à l'évaluation d'un large évantail de fonctions physiques, sociales et affectives. Ainsi, la plupart des mesures de la qualité de vie recouvrent les dimensions ou les caractéristiques suivantes :

a) la fonction sensorielle (la parole, la vue et l'ouïe),
b) la mobilité et la fonction physique,
c) la fonction affective,
d) la fonction cognitive,

e) l'autonomie dans les activités de la vie quotidienne, et

f) le niveau de douleur et d'inconfort.

La qualité de vie globale est modifiée par toute une gamme de facteurs qui ne relèvent pas de l'état de santé. La situation socio-économique en est un exemple frappant. Dans le cadre des études cliniques, toutefois, on postule que ces autres facteurs sont contrebalancés par la méthodologie expérimentale et que l'on peut, par conséquent, les ignorer lorsqu'il s'agit de comparer les effets de différents traitements sur la qualité de vie. La section qui suit porte principalement sur la mesure de la qualité de vie liée à la santé. Plus particulièrement, on s'attardera à la façon d'utiliser ces mesures pour évaluer les risques et les avantages des drogues nouvelles.

Lorsque l'on soumet une drogue nouvelle à un essai, la mesure de la qualité de vie liée à la santé (QVLS) prend de nombreuses dimensions. Les avantages attribués au médicament doivent être pesés à la lumière des effets secondaires qu'il engendre. La qualité de vie (et le respect et l'efficacité de la médication) n'est pas influencée que par les effets secondaires. D'autres propriétés du médicament entrent en ligne de compte : par exemple, de façon générale, les préparations à prendre par la bouche sont préférées aux préparations injectables, les injections hypodermiques à faire soi-même ont la préférence sur les administrations intraveineuses de médicaments faites par le médecin, les médicaments qui ont une saveur sont préférés à ceux au goût insipide, on aime mieux recevoir une dose d'un médicament à longue durée d'action que plusieurs doses d'un médicament à courte durée d'action, les médicaments qui sont faciles à conserver, à préparer et à transporter obtiennent la préférence sur ceux qui ne le sont pas, et les médicaments qui coûtent moins cher se vendent mieux que les médicaments dont le prix est élevé.

Les différentes approches de la mesure de la qualité de vie par rapport à la santé sont étudiées au chapitre 5.2. L'une des principales est celle de l'utilité, dont les principales caractéristiques sont étudiées au chapitre 5.3. La mesure de l'utilité est décrite au chapitre 5.4. Les avantages et les inconvénients du recours à la qualité de vie en fonction de l'utilité dans l'évaluation des drogues nouvelles, en vue de leur autorisation, sont présentés aux chapitres 6.1 et 6.2. Enfin, le chapitre 6.3 illustre l'évaluation de la qualité de vie fondée sur l'utilité d'une drogue nouvelle. Ce produit est un sel d'or oral (auranofine) utilisé pour le traitement des patients arthritiques.

5.2 Méthodes de mesure de la qualité de vie

On a assisté à la naissance d'une grande variété de méthodes d'évaluation de la qualité de vie. Une telle diversité est à la fois inévitable et appropriée. La qualité de vie est en effet un concept vaste et multi-dimensionnel, et des mesures différentes permettent d'en considérer les différents aspects de vie. Par ailleurs, les investigateurs ont une grande variété d'objectifs, ce qui peut demander une batterie d'instruments et de méthodes convenant à différentes situations et questions de recherche. À présent, on peut dire qu'il n'existe pas de règle d'or pour la mesure de la qualité de vie, et il y a peu de chance qu'il en apparaisse une un jour.

Il y a différentes façons de classer les mesures de la qualité de vie. Un schéma de classification est présenté au tableau 2 (page 49). Des *instruments spécifiques* sont mis au point, pour s'appliquer à des situations cliniques particulières. Généralement, ces mesures visent à évaluer le fardeau de la maladie (ou le problème de santé) et de son traitement pour la qualité de vie. En y incluant que les éléments qui sont les plus importants dans la situation clinique particulière, ces mesures peuvent être établies de façon à comprendre une large gamme d'effets (p. ex., fonctionnement physique, fonctionnement émotionnel, caractéristiques qui sont d'un intérêt spécial pour les patients individuels) sans imposer un trop lourd fardeau aux répondants. Ces mesures ont aussi été établies de façon à être le plus adaptable possible au changement. Un bon exemple d'instrument spécifique est le Questionnaire des maladies respiratoires chroniques élaboré par Guyatt et al (1987).

Les *instruments génériques* sont conçus pour s'appliquer à une grande variété de situations. Ces mesures comprennent un certain nombre de dimensions : fonctionnement physique, fonctionnement social et émotionnel, douleurs, soins personnels et autres activités. Il existe deux catégories principales de mesures génériques, soit les *profils de santé* et les *mesures d'utilité pour la santé* (Guyatt, van Zanten, Feeny et Patrick, 1988; et Feeny, Labelle, Torrance en préparation).

a) Les profils de santé et les mesures d'utilité pour la santé (Guyatt, van Zanten, Feeny et Patrick, 1988; et Feeny, Labelle, Torrance, en préparation). Dans les profils de santé, on obtient des scores pour chaque domaine approprié de qualité de vie par rapport à la santé. Le profil peut ou non donner un seul score résumé. Le profil d'impact de maladie (Bergner, 1981) est un bon exemple d'une méthode de profil de santé qui donne un seul score résumé, à la différence du profil de santé Nottingham.

b) La deuxième catégorie de mesure générique, la méthode de l'utilité, est analysée dans les sections suivantes. La prochaine section passera brièvement en revue les notions fondamentales de la méthode de l'utilité pour évaluer la qualité de vie.

5.3 Définition de l'« utilité »

L'*utilité*, notion utilisée en économique et en analyse décisionnelle, se rapporte à un degré de satisfaction ou de plaisir. En général, les économistes ne tentent pas de mesurer directement l'utilité (préférences des consommateurs pour diverses solutions de consommation). Dans le contexte de l'analyse du comportement dans la plupart des situations de marché, l'observation du comportement des consommateurs permet de déduire leurs préférences. C'est pourquoi la discipline de l'économie dispose d'un ensemble bien au point d'outils statistiques destinés à l'analyse des données tirées de l'expérience, mais la gamme d'outils dont elle dispose pour la mesure directe de l'utilité est beaucoup moins étendue.

À l'opposé des économistes, les analystes de décisions effectuent régulièrement la mesure des préférences des décideurs. Les mesures sont fondées directement sur les axiomes de la théorie de l'utilité qui ont été postulés pour la première fois par von Neumann et Morgenstern (1944). Si cette méthode continue de susciter des controverses (Samuelson et Zeckhauser, 1988), ses axiomes de base sont depuis plus de 40 ans le modèle normatif dominant de la prise de décision rationnelle quand les résultats sont incertains. La méthode classique (règle d'or) de mesure de l'utilité consiste à utiliser la technique de pari type qui vient directement des axiomes de base. Les utilités mesurées de cette façon sont justement nommées « utilités de von Neumann-Morgenstern » pour éviter toute confusion. C'est la définition de l'utilité donnée par von Neumann et Morgenstern qui sera utilisée dans le présent rapport.

5.4 Mesure de l'utilité

Pour la mesure de l'utilité, on a mis au point une petite famille de méthodes qui sont fiables et valides. Une brève description des étapes pratiques de la méthode est donnée, mais on trouvera des descriptions plus détaillées dans Torrance, 1986; Drummond, Stoddart et Torrance, 1987; ou Feeny et Torrance, 1989. Par exemple, on décrit un certain nombre de scénarios d'état de santé, comprenant peut-être l'état de santé courant du patient défini subjectivement. Ces scénarios peuvent couvrir une large gamme de dimensions, comprenant le fonctionnement physique, le fonctionnement sensoriel, le fonctionnement cognitif, le bien-être émotionnel, la douleur, les soins personnels, et toutes les caractéristiques spéciales de la maladie ou de son traitement (un exemple est fourni au tableau 3, page 50). On demande alors aux sujets de classer les états de santé par ordre de priorité, souvent à l'aide d'une échelle visuelle analogue marquée de 0 à 100, connue sous le nom de thermomètre d'états de santé (voir figure 3, page 45). Les mesures données par le thermomètre d'états de santé transmettent à l'investigateur une nette indication du classement ordinal des états de santé (des divers résultats de traitement) et une certaine information sur l'intensité de ses préférences. Si le thermomètre des états de santé semble donner des valeurs d'échelle à

intervalles, les observations rassemblées jusqu'à ce jour indiquent que s'il donne une bonne indication des classements par ordre de préférence, il ne fournit pas directement des scores d'utilité cardinaux.

La mesure des scores cardinaux pour l'utilité des états de santé s'effectue au moyen d'une des deux techniques suivantes : le « pari type » ou la « durée de vie équivalente ». Dans la méthode du pari type, le sujet se voit proposer un choix entre deux solutions : le choix A consiste à prendre l'état de santé en question avec certitude, tandis que le choix B consiste à prendre un pari sur un traitement qui a la possibilité d'un résultat meilleur ou d'un résultat moins bon. La figure 4 (page 46) représente une application de cette méthode à la mesure de l'utilité pour un état chronique appelé état A. Ici, le choix B conduit à une santé parfaite avec une probabilité p, et à une mort immédiate avec la probabilité $1-p$. On fait ensuite varier la probabilité p jusqu'à ce que le sujet reste indifférent entre les choix A et B. Afin de voir qu'il doit exister une et une seule probabilité d'indifférence, considérons le cas où l'état A est, aux yeux du sujet, préférable à la mort, mais moins bon que la santé parfaite. Quand p est égal à 1,0, le sujet préférera le choix B parce qu'il conduit avec certitude à une santé parfaite. Mais si p est égal à 0,0, le sujet préférera le choix A parce que le choix B conduit avec certitude à une mort immédiate. Quelque part entre $p = 1,0$ et $p = 0,0$, il existe une valeur de p, comme sous le nom de probabilité de l'indifférence, sur laquelle le sujet doit fixer son choix.

L'échelle d'utilité pour les états de santé se définit par une utilité de 1,0 pour la santé parfaite, et 0,0 pour la mort. Ensuite, les axiomes de la théorie de l'utilité spécifient que l'utilité pour l'état A est simplement égale à la probabilité de l'indifférence. Par exemple, si un sujet est indifférent entre les choix A et B quand $p = 0,75$, l'utilité pour l'état A est égale à 0,75. Il est à noter que plus la probabilité de l'indifférence est faible, plus le risque de mort que le sujet est prêt à considérer est élevé, et ainsi plus l'utilité pour l'état A est faible.

Comme certains répondants éprouvaient des difficultés à comprendre les probabilités, une autre technique, appelée celle de la durée de vie équivalente, a été mise au point. Bien que cette technique ne soit pas fondée directement sur les axiomes de la théorie de l'utilité, il a été observé de façon empirique qu'elle fournissait une approximation serrée (Torrance, 1976). L'application de cette méthode à la mesure de l'utilité pour un état chronique est représentée par la figure 5 (page 47). Dans cette technique, le sujet se voit offrir également deux choix. Le choix A est le même que dans le pari type : l'état de santé A pour le reste de la vie, c'est-à-dire un nombre d'années à vivre égal à t. Le choix B consiste à choisir une santé parfaite pendant une période plus courte, x années, suivie par la mort. On fait varier le temps x jusqu'à ce que le sujet soit indifférent entre la plus courte période en parfaite santé et la plus longue période en moins bonne santé. La valeur de l'utilité pour l'état de santé A s'exprime par le rapport x/t. Par exemple, si le sujet est indifférent entre 40 ans à vivre avec l'état de santé A et 30 ans à vivre en bonne santé, la valeur de l'utilité pour l'état de santé A est égale à 30/40, ou 0,75.

Les techniques du pari type, de la durée de vie équivalente, ainsi que les méthodes qui en dérivent, permettent d'obtenir de façon fiable et valide des scores d'utilité pour des états de santé. L'usage préalable du thermomètre des états de santé semble aider les répondants à effectuer leur introspection concernant leurs préférences pour les états de santé en cours d'évaluation. Les états de santé à évaluer peuvent comprendre ceux que le répondant a déjà vécus, ceux qu'il vit ou des états qu'il n'a jamais éprouvés (états hypothétiques). À cet égard, la méthode de l'utilité diffère des autres mesures de qualité de vie. La méthode de l'utilité, quand elle comprend des états de santé hypothétiques, permet à l'investigateur d'obtenir des renseignements importants de tous les patients, sur la façon dont ils réagiraient s'ils faisaient l'expérience de certaines des perspectives peu fréquentes. Les résultats obtenus jusqu'à ce jour démontrent que les évaluations de personnes connaissant un état de santé donné et par d'autres personnes pour lesquelles cet état de santé est hypothétique, ne diffèrent généralement pas considérablement (Sackett et Torrance, 1978; Wolfson et al, 1982). S'il existe une différence systématique, c'est que l'état de santé est coté légèrement plus haut par les personnes qui connaissent cet état de santé que par les autres; mais cette différence n'est encore pas concluante. Ainsi, la qualité de vie pour des perspectives rares ou peu fréquentes peut être évaluée en même temps que celle ayant trait à des perspectives fréquentes, en combinant des états de santé existants et hypothétiques dans le jeu de perspectives à évaluer.

6. EXPLORATION DE LA MESURE DE LA QUALITÉ DE VIE FONDÉE SUR L'UTILITÉ DANS L'ÉVALUATION DES DROGUES NOUVELLES EN VUE DE LEUR AUTORISATION

La question du recours aux mesures d'utilité de la qualité de vie dépend en partie des caractéristiques de base et des avantages de la méthode.

6.1 Avantages des mesures de l'utilité

L'utilisation de la méthode de l'utilité dans la mesure de la qualité de vie présente les avantages suivants : on peut la généraliser; elle est globale; elle permet d'intégrer les effets de mortalité et de morbidité; elle permet d'intégrer les risques et les avantages; elle permet de déterminer les préférences à la fois pour les situations hypothétiques et les situations réelles; elle permet de représenter des points de vue multiples; elle permet d'utiliser différentes sortes d'évaluateurs pour les états de santé; elle permet d'incorporer les préférences des sujets pour ce qui est de la durée de vie et des risques dans leurs évaluations d'états de santé; elle repose sur des bases théoriques solides; et elle est compatible avec l'évaluation économique (voir tableau 4, page 51).

6.1.1 Possibilité de généralisation

Les scores obtenus par la méthode de l'utilité sont représentés sur une échelle généralisable allant de 0,0 (mort) à 1,0 (parfaite santé). Cette échelle est tout à fait générale et peut être utilisée pour mesurer l'impact de tous les médicaments, quelle que soit la maladie ou le type de patient.

6.1.2 Portée

Les mesures de l'utilité sont parmi les mesures de la qualité de vie qui ont la plus grande portée. La description des états de santé comprend de façon générale des renseignements sur les niveaux de fonctionnement sensoriel, physique, socio-émotionnel et cognitif, et de soins personnels, de même que sur les symptômes ou problèmes de santé, effets secondaires et niveaux de douleur. De cette façon, l'évaluateur d'un état de santé particulier intègre toutes les caractéristiques appropriées de la qualité de vie par rapport à la santé dans la mesure du bien-être général fondé sur l'utilité.

6.1.3 Intégration des effets de morbidité et de mortalité

La méthode de l'utilité permet à l'analyste de combiner les effets de mortalité et de morbidité en une seule mesure sommaire de l'amélioration de la santé. Normalement, on y parvient par la méthode de l'utilité combinée à celle des années de vie ajustées en fonction de la qualité, en utilisant des utilités comme ajustements de la qualité (Weinstein et Stason, 1977; Drummond, Stoddart et Torrance, 1987; et Torrance et Feeny, 1990.) On peut également opérer directement avec la méthode de l'utilité en définissant et en mesurant une fonction d'utilité d'après les deux caractéristiques principales, soit la durée de vie et la qualité de vie (Pliskin et al, 1980).

6.1.4 Saisie de deux dimensions de risque

L'utilité attendue permet de capter deux dimensions de risques : (1) la « probabilité » d'une perspective et (2) le « degré de préférence » pour cette perspective. D'autres méthodes d'analyse risques-avantages ne permettent pas de capter les deux dimensions en un seul nombre.

6.1.5 Intégration des risques et avantages

La méthode de l'utilité mesure tous les impacts, positifs ou négatifs, d'une drogue nouvelle de prescription, avec les mêmes unités communes. De cette façon, elle donne la possibilité d'intégrer les risques et les avantages, et de déterminer si le bilan risques-avantages est favorable ou défavorable par le même patient ou l'ensemble des patients.

6.1.6 Mesure des préférences pour des états de santé hypothétiques et existants

Un sixième avantage est que la même personne peut évaluer à la fois des états de santé hypothétiques et existants. La plupart des méthodes de mesure de la qualité de vie et des tests psychosociaux exigent que le sujet vive l'état de santé au moment où l'évaluation est effectuée. On ne peut donc obtenir des évaluations pour toutes les perspectives possibles. Le problème est particulièrement aigu pour les états de santé qui sont peu fréquents. Dans la méthode de l'utilité, des données valables peuvent être obtenues en évaluant des états de santé hypothétiques relativement à des états de santé existants, pour lesquels des scores sont aussi obtenus.

6.1.7 Représentation de points de vue multiples

Une septième caractéristique de la méthode de l'utilité est qu'on peut utiliser une grande variété d'évaluateurs pour les mêmes états de santé : les patients, leurs parents et des membres du public. Les comparaisons entre évaluations peuvent être importantes et riches en informations. Cet outil permet d'effectuer l'évaluation d'une technologie en partant de plusieurs points de vue, dont le plus étroit est celui du patient, et le plus large celui de la société.

6.1.8 Incorporation des préférences de temps et de risque

Une huitième caractéristique de la méthode de l'utilité est qu'elle permet de capter à la fois la préférence du répondant en matière de temps (durée de vie et pari type) et ses attitudes à l'égard du risque (pari type). Étant donné que l'incertitude de la perspective future est implicite dans la plupart des situations cliniques, la méthode de l'utilité est un modèle particulièrement valable.

6.1.9 Bases théoriques rigoureuses

Un avantage important de la méthode de l'utilité est qu'elle possède des bases théoriques rigoureuses et explicites représentées par la théorie de l'utilité de von Neumann et Morgenstern.

6.1.10 Intégration aux mesures d'évaluation économique

Un autre avantage de la méthode de mesure de l'utilité est qu'elle est peut-être intégrée à des mesures d'évaluation économique. Elle permet donc de comparer des perspectives parmi des interventions diverses en matière de soins de santé, au moyen de l'analyse coût-utilité et de ses variantes. Les conceptions de base de l'évaluation de l'utilité sont en accord avec les bases théoriques de l'analyse coûts-avantages en économie.

6.1.11 Intégration aux essais cliniques

Les mesures de qualité de vie fondées sur l'utilité peuvent être intégrées dans des essais cliniques prospectifs. La randomisation et les essais en aveugle permettent d'obtenir une mesure valide et non biaisée. De plus, en associant les investigations cliniques et les mesures de qualité de vie dans une seule étude, on obtient des résultats qui sont tous disponibles d'une façon opportune pour jouer un rôle dans une politique clinique et de soins de santé.

6.1.12 Utilité pour les médecins prescripteurs

Les résultats relatifs à la qualité de vie fondés sur l'utilité qui sont obtenus à partir des essais cliniques en vue d'une décision d'autorisation peuvent être utiles plus tard au médecin prescripteur. Par exemple, les résultats peuvent indiquer, par type de patient, le bilan risques-avantages escompté et sa distribution. Il serait possible de résumer et de présenter ces informations d'une façon qui, plus tard, aiderait le médecin prescripteur ou le patient à estimer le bilan risques-avantages qui se présente à ce dernier.

En somme, la méthode de l'utilité permet de faire des évaluations globales de la qualité de vie par rapport à la santé qui s'étendent sur une vaste gamme d'applications. Le résultat qu'elle donne est un simple score sur une échelle généralisable allant de 0,0 (mort) à 1,0 (parfaite santé). L'information sur les attitudes face au risque, les préférences pour ce qui est du temps à vivre et les choix parmi les avantages du traitement et ses effets secondaires est reflétée dans les scores d'utilité. L'analyse de l'utilité peut aussi être facilement combinée avec des mesures financières (en dollars) des coûts et avantages des différents choix de soins de santé. Enfin, les deux mesures peuvent être incorporées dans des essais cliniques randomisés.

6.2 Inconvénients de la méthode de l'utilité dans la mesure de la qualité de vie

L'utilisation de la méthode de l'utilité pour mesurer la qualité de vie en relation avec la santé comporte certains inconvénients (tableau 4, page 51).

a) D'abord, si les mesures des analyses d'utilité sont fiables et valides, elles ne sont pas extrêmement précises. Pour les scores individuels, l'écart-type est d'environ 0,13 (Torrance, 1976).

b) Deuxièmement, pour obtenir des mesures fiables, on doit pratiquer des interviews qui prennent beaucoup de temps et doivent être conduites par du personnel professionnel.

c) Troisièmement, les scores d'utilité peuvent ne pas être compris rapidement ou interprétés facilement.

d) Quatrièmement, si les mesures de l'utilité ont été obtenues avec succès d'une grande variété de types de répondants, elles ne peuvent être appliquées de façon universelle. La complexité des méthodes semble demander de réelles connaissances linguistiques pour déchiffrer le langage dans lequel les documents sont présentés. Il existe également des restrictions liées à l'âge et à la capacité cognitive.

e) Finalement, tandis qu'un des principaux avantages de la mesure de l'utilité est son score unique et généralisable basé sur une intégration subjective par la personne, cette méthode ne révèle pas la raison pour laquelle une valeur d'utilité particulière peut être élevée ou faible. Il peut être nécessaire d'inclure dans l'essai d'autres mesures de qualité de vie pour permettre de telles déterminations.

6.3 Incorporation de la mesure de la qualité de vie dans un essai de médicament : exemple

Récemment, des mesures de qualité de vie ont été incorporées dans un essai international multi-centrique portant sur un nouveau produit pharmacologique, l'auranofine, destiné au traitement de l'arthrite (Bombardier et al, 1986). Dans cet essai, les mesures de qualité de vie étaient appropriées parce que la perspective recherchée était une amélioration de la qualité de vie, et non une réduction de la mortalité. Pour ces patients, la qualité de vie ne serait pas affectée seulement par la douleur et les autres effets fâcheux de la maladie, mais également par les effets secondaires du traitement lui-même. Ces effets secondaires comprennent, entre autres : dermatite, stomatite, thrombocytopénie, leucopénie et protéinurie, diarrhée, selles liquides et crampes abdominales. Cependant, les accidents de toxicité aiguë sont peu fréquents, et les cas de décès associés à la médication sont rares.

L'étude avait pour but de comparer l'auranofine (3 mg deux fois par jour) à un placebo, comme traitement supplémentaire administré à des patients atteints de polyarthrite rheumatoïde et stabilisés au moyen d'un traitement de base prudent comprenant le repos, la physiothérapie, et l'administration d'anti-inflammatoires non-stéroïdiens, de stéroïdiens oraux et d'analgésiques. Un total de 303 patients, dans 14 centres, ont été suivis pendant 6 mois dans le cadre d'une étude randomisée à double insu. Une grande variété de perspectives de résultats du traitement ont été employés comme mesures (voir tableau 5, page 52 et 6, page 53). Jusqu'à ce jour, l'étude fournit une des comparaisons directes les plus serrées des mesures possibles dans un essai clinique. On a utilisé dans l'essai un « instrument de mesure d'utilité patient »*, élaboré spécialement pour l'étude. Les autres résultats mesurés étaient des mesures cliniques, des mesures d'état fonctionnel, des mesures de la douleur, des mesures d'impression globale et autres.

* « patient-utility measurements (PUMS) »

L'instrument de mesure d'utilité patient est basé sur des techniques de mesure de l'utilité bien établies. Il est destiné à quantifier, au moyen d'une interview du patient après traitement, le changement dans la qualité de vie apporté par un traitement. L'instrument consiste en un manuel et en un questionnaire destinés à l'utilisateur, complété par des aides visuelles perfectionnées sous la forme d'un paquet de cartes, le tout étant conçu pour être utilisé facilement, à la fois par l'intervieweur et par le patient. L'instrument est conçu pour être applicable à des patients adultes de tout âge, pour neutraliser les biais déformants potentiels, et pour contrer les biais de mesure connus au moyen d'une stratégie dite de ping-pong convergent.

L'instrument de mesure comportait trois parties.

- La **partie 1** mesurait le changement dans la qualité de vie entre la période de départ avant l'essai et le moment de l'interview. Ce changement était mesuré de trois façons : par une échelle d'évaluation, une loterie et un choix de temps à vivre modifié. L'échelle d'évaluation était une variante du thermomètre des états de santé (figure 3, page 45) adaptée pour mesurer un changement plutôt que pour donner un score absolu. De même, la loterie et le choix de temps à vivre modifié donnaient des scores qui étaient des adaptations du pari type (figure 4, page 46) et du choix de temps à vivre (figure 5, page 47), respectivement.
- La **partie 2** mesurait la qualité de vie absolue au départ et au moment de l'interview, au moyen du choix de temps à vivre conventionnel (figure 5, page 47).
- La **partie 3** mesurait l'« aversion au risque »; plus précisément, elle mesurait la concavité/convexité de la fonction d'utilité pour une durée de vie supplémentaire en bonne santé. Une adaptation du pari type était utilisée dans cette partie (figure 4, page 44).

L'instrument de mesure d'utilité-patient a été appliqué à 239 patients dans cet essai (122 : auranofine, 117 : placebo), par des intervieweurs dans chacun de 14 centres différents. La durée moyenne des interviews était de 32 minutes (limites 8 et 82 minutes). Ces interviews étaient enregistrées; une centaine de ces interviews ont été analysées en profondeur, et les chercheurs n'ont rencontré aucune difficulté importante dans l'une ou l'autre des parties de l'instrument.

L'instrument de mesure d'utilité patient a donné des résultats indiquant que l'état des patients auranofine et placebo s'améliorait au cours de l'essai, mais que l'état des patients auranofine s'améliorait encore plus. De façon précise, l'amélioration des patients placebo était représentée par une élévation du niveau de qualité de vie, qui passait de 0,629 à 0,728, soit une augmentation de 0,099, ou 16 %. Chez les patients auranofine, le niveau de vie passait de 0,571 à 0,781, soit une augmentation de 0,209, ou 37 %. La différence était significative à $p = 0,002$.

Une grande variété de résultats ont été mesurés dans l'essai, et toutes ces mesures n'ont pas été également sensibles (voir tableaux 5, page 52 et 6, page 53). L'auranofine a été trouvée supérieure au placebo par un indice composite de mesures de résultats cliniques (p = 0,003), de mesures d'état fonctionnel (p = 0,001), de mesures d'impression globale (p = 0,007), ainsi que par l'indice de qualité de bien-être de Bush (p = 0,005). Une mesure composite de la douleur a montré une amélioration (p = 0,021), moins inférieure au niveau de signification prévue dans l'essai : p = 0,0125. Un pari type conventionnel entre la bonne santé et la mort à la fin du traitement a donné des résultats numériques analogues à ceux de l'instrument de mesure d'utilité patient (PUMS) (utilité auranofine égale à 0,77, utilité placebo 0,70, comparativement aux valeurs PUMS de 0,78 et 0,73), mais n'a pas tout à fait permis d'atteindre la signification statistique (p = 0,07), probablement parce que cette méthode n'était pas, à la différence du PUMS, destinée à mesurer des variations.

Pratiquement aucune différence n'a pu être détectée par la méthode de disposition à payer, une technique relativement récente qui présentait, semble-t-il, de graves difficultés de mesure (Klarman, 1982; Thompson et al, 1982; Thompson et al, 1984). Essentiellement, la méthode de disposition à payer tend à déterminer la somme en dollars qu'une personne serait disposée à payer pour se débarrasser de la maladie. Si le tableau 6 (page 53), tiré de l'article original (Bombardier et al, 1986) classe la disposition à payer comme mesure d'utilité, ce n'est pas strictement exact. La disposition à payer donne une mesure en dollars, non une mesure de l'utilité; cependant, cette mesure en dollars est conçue pour se substituer au besoin des mesures d'utilité dans les évaluations économiques.

De même, l'échelle de dépression du N*ational Institute of Mental Health* n'a pas montré de différence, tandis que le *Rand General Health Perceptions Questionnaire* ne révélait qu'une différence modeste, non significative. Aucun de ces instruments n'est une mesure de l'utilité.

L'instrument de mesure d'utilité patient s'est révélé l'un des instruments de prévision les plus sensibles dans l'essai, tel que l'indique le niveau de signification statistique qu'il permet d'atteindre. En fait, sa sensibilité n'a été dépassée que par deux mesures, le score composite fonction-état de santé, et l'échelle catégorielle de l'arthrite. De plus, les scores composites ont diminué la puissance statistique, par comparaison avec les scores des instruments individuels. Par exemple, chacune des trois composantes individuelles du score composite fonction-état de santé a été moins sensible que le PUMS. Enfin, à la différence des autres mesures, le PUMS donne également une évaluation de l'impact global sur la qualité de vie.

C'est ainsi que si l'on compare directement les 25 résultats individuels mesurés dans l'essai, la sensibilité du PUMS n'était dépassée que par celle de l'échelle catégorielle de l'arthrite. Dans cette dernière, les patients indiquent essentiellement que leur condition s'est améliorée. Dans le PUMS, par contre, ils font équivaloir cette amélioration a un impact globale sur leur qualité de vie en rapport avec la santé.

Nota : Il ne faut cependant pas oublier que le PUMS est spécifique à une étude et qu'il ne peut être utilisé directement dans d'autres études sans subir de modifications. Néanmoins, son succès dans cette étude suggère que des instruments similaires peuvent être élaborés pour d'autres essais cliniques et que la méthode de l'utilité est appropriée à chaque fois qu'on désire une évaluation globale des perspectives.

Un des principaux avantages du PUMS est sa simplicité d'utilisation. Il est essentiellement constitué d'un paquet de cartes qui ne nécessite presque pas de directives et qui est complété par un jeu précis d'instructions pour les interviews. Avec une formation approfondie et un bon contrôle de qualité, les assistants de recherche de 14 centres répartis dans deux pays ont été capables d'utiliser l'instrument de mesure d'une façon homogène, et sans problèmes.

L'exemple de l'auranofine illustre le fait que la méthode de mesure de l'utilité peut être incorporée de façon viable dans des essais cliniques et utilisée pour évaluer la qualité de vie pour diverses perspectives. L'expérience indique que si l'on élabore et met en jeu avec soin des instruments de mesure d'utilité spécifiques à une étude, ils sont fiables, valides et sensibles.

7. RÉSUMÉ ET CONCLUSIONS

La réduction du fardeau de la maladie et l'amélioration de la qualité de vie sont parmi les objectifs de traitement fixés pour un grand nombre de nouveaux produits pharmaceutiques. De plus, un grand nombre, sinon la plupart, des médicaments comportent des effets secondaires et/ou des réactions indésirables qui ont un impact sur la qualité de vie des personnes qui reçoivent ces médicaments. Étant donné que la qualité de vie est une perspective clé, il est important et désirable de pouvoir la mesurer directement dans des essais thérapeutiques.

Le développement rapide au cours des dernières années de méthodologies améliorées permettant de mesurer la qualité de vie la santé rend maintenant possible la mesure directe des effets sur la qualité de vie dans les essais cliniques. Un certain nombre d'instruments de mesure spécifiques ont été utilisés avec succès dans les essais thérapeutiques pour une grande variété de maladies : les pneumopathies chroniques, la chimiothérapie pour le traitement du cancer du sein, l'arthrite et l'hypertension. Un certain nombre d'instruments de mesure spécifiques, comme le profil de l'impact de la maladie, ont également été utilisés avec succès dans une variété de contextes cliniques. Finalement, dans quelques essais, on a utilisé avec succès des mesures d'utilité pour la qualité de vie par rapport à la santé. D'une façon générale, dans ces applications, les mesures de qualité de vie ont complété et élargi les mesures cliniques traditionnelles des perspectives. Les mesures de la qualité de vie ont permis d'obtenir de nouvelles informations venant s'ajouter à celles provenant d'autres mesures, et ont eu l'avantage supplémentaire de mesurer directement la perspective qui, souvent, est celle d'une importance fondamentale : la qualité de vie du patient. S'il est vrai que des progrès supplémentaires dans le domaine du développement méthodologique sont encore nécessaires et qu'il n'existe pas encore de règle d'or, l'industrie pharmaceutique et les autorités de réglementation comme la DGPS devraient continuer de développer l'expérience et l'expertise acquises dans le domaine de ces méthodologies. Il est maintenant raisonnable de prendre en compte les résultats obtenus avec ces instruments de mesure dans les demandes d'autorisation de drogues nouvelles. La qualité de vie en relation avec la santé est une mesure des résultats tout à fait valide.

L'utilisation des mesures de la qualité de vie ajoute des informations importantes à celles dont on disposait déjà sur les effets des traitements. L'élaboration d'instruments fiables permettant de mesurer objectivement un phénomène subjectif, comme les changements dans la qualité de vie, ajoute une importante dimension à l'analyse risques-avantages. Des instruments de mesure de la qualité de vie spécifiques aux maladies et les mesures génériques comme le profil de santé fournissent une information sur les variations de l'état de santé telles qu'elles sont perçues par les patients, et sur l'importance que ceux-ci leur attribuent. Souvent, d'ailleurs, ces instruments de mesure permettent d'observer l'impact des effets secondaires sur la qualité de vie. Un certain nombre de ces instruments se sont révélés utiles dans l'évaluation d'un nombre réduit mais diversifié de médicaments.

L'utilisation de toute mesure fiable, valide et sensible de la qualité de vie par rapport à la santé apporte une information utile à l'analyse risques-avantages. Une méthode, cependant, assure un cadre plus global et plus rigoureux qui permet d'évaluer les choix entre les effets positifs et négatifs d'un traitement : la méthode de l'utilité, qui donne un score sommaire unique qui reflète l'impact net sur la qualité de vie. Les patients évaluent subjectivement, en fonction de leurs préférences, le niveau d'amélioration de leur état de santé, comparativement au fardeau des effets secondaires résultant d'un traitement médicamenteux. En associant les états de santé hypothétiques et existants, on peut également représenter, dans la même évaluation, les répercussions d'effets secondaires peu fréquents mais importants. De plus, la méthode de l'utilité intègre les effets de mortalité et de morbidité, donnant ainsi une structure très globale permettant d'évaluer les effets d'un traitement médicamenteux. Cette méthode est un instrument de mesure extrêmement approprié et intéressant qui pousse plus loin l'analyse risques-avantages traditionnelle.

8. EN CONCLUSION

Les méthodologies de mesure de la qualité de vie en relation avec la santé, y compris la méthode de l'utilité, sont maintenant suffisamment bien développées pour qu'il soit scientifiquement approprié de les utiliser à une échelle modeste dans les essais thérapeutiques. Les mesures de la qualité de vie ajoutent une information significative et utile à celles qui sont traditionnellement soumises en vue d'une autorisation de médicament. La méthode fondée sur l'utilité employée pour mesurer la qualité de vie est particulièrement puissante, intéressante et digne de considération.

9. BIBLIOGRAPHIE

ALTSHULER, B. A Bayesian Approach to Assessing Population Risks from Environmental Carcinogens. *Proceedings SIMS Conference on Environmental Health*, 1976, Alta, Utah.

BANTA H.D., BEHNEY C.J. et WILLEMS J.S. *Toward Rational Technology in Health Care*. New York: Springer Publishing Company, Inc. 1981, p. 93.

BERGNER, M., BOBBITT, R.A., CARTER, W.B. et GILSON, B.S. The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. *Medical Care*, 1981, 19: 787–805.

BOMBARDIER, C., WARE, J., RUSSELL, IL., LARSON, M., CHALMERS, A. et READ, J.L. Auranofin Therapy and Quality of life in Patients with Rheumatoid Arthritis. *Am. J. Med.*, 1986, 81: 565–578.

BUXTON, M.J. et DRUMMOND, M.F. Quality of Life Measurement and the Regulation of Medicines. (version non publiée) 1988.

COMMITTEE OF PRINCIPAL INVESTIGATORS. W.H.O. Cooperative Trial on Primary Prevention of Ischemic Heart Disease Using Clofibrate to Lower Serum Cholesterol: Mortality Follow-up. *Lancet*, 23 août 1980: 379–385.

DEPARTMENT OF CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS, McMaster University, Hamilton. 1984. How to Read Clinical Journals: VII. To Understand an Economic Evaluation (Part A). *Can. Med. Assoc. J.*, 130: 1428–1433.

DEPARTMENT OF CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS, McMaster University, Hamilton. How to Read Clinical Journals: IV. To Determine Etiology or Causation. *Can. Med. Assoc. J.*, 1981, 124: 985–990.

DRUMMOND, M.F., STODDART, G.L. et TORRANCE, G.W. *Methods for the Economic Evaluation of Health Care Programs*. Oxford: Oxford University Press. 1987.

FEENY, D. Can We Use Quality-of-Life Measures for Economic Evaluation? *Quality of Life in Cardiovascular Care* (à paraître). 1988.

FEENY, BARR, FURLONG, TORRANCE, ROSENBAUM, WEITZMAN et HORSMAN (en préparation). Economic Evaluation of Treatment for Childhood Cancer: Costs, Benefits, and the Quality of Life.

FEENY, D., GUYATT, G. et TUGWELL, P. (éd.). Health Care Technology: *Effectiveness, Efficiency and Public Policy*. The Institute for Research on Public Policy, Montréal. 1986.

FEENY, D., LABELLE, R. et TORRANCE, G.W. Chapter 6: Integrating Economic Evaluations and Quality-of-Life Assessments. *Quality of Life Studies* Bert Spilker (éd.). Raven Press, New York. 1990 (sans presse).

FEENY, D. et TORRANCE, G.W. Incorporating Utility-Based Quality-of-Life Assessment Measures in Clinical Trials: Two Examples. *Medical Care*, 1989, 27(3):5190–5204.

FEINSTEIN, A.R. Clinical Biostatistics. IX. How Do We Measure "Safety and Efficacy" *Clin. Pharm. S. Ther.*, 12(3): 544–558.

FINEBERG, H.V. Decision Trees: Construction, Uses and Limits. *Bull. Cancer*, 1980, 67: 395–404.

GILMAN, A., GOODMAN, L.S., RALL, T.W. et FREID, M. (éd.). *The Pharmacological Basis of Therapeutics* (7e éd.). Macmillan Publishing Co., London. 1985.

GUYATT, G.H., TOWNSEND, M., BERMAN, L.B. et PUGSLEY, S. O. Quality of Life in Patients with Chronic Airflow Limitations. *Br. J. Dis. Chest*, 1987. 81: 45–54.

GUYATT, VAN ZANTEN, FEENY et PATRICK (en préparation). Measuring Quality of Life in Clinical Trials. Manuscript, 1988.

HAYNES, R.B., MUKHERJEE, J., SACKETT, D.L. et al. Functional Status Following Medical or Surgical Treatment for Cerebral Ischemia. *J. Am. Med. Assoc.*, 1987, 257(15): 2043–2046.

JONES, D.R. et AKEHURST, R.L. Risk Assessment: An Outline. *The Institute of Mathematics and its Applications*, 1980, 16: 252–258.

KLARMAN, H.E. The Road to Cost-effectiveness Analysis. *Milbank Memorial Fund Quarterly/Health and Society*, 1982, 60(4): 585–603.

KNIGHT, F. Risk, *Uncertainty, and Profit*. Houghton Miffin Co., Boston. 1921.

LAUPACIS, A., SACKETT, D.L. et ROBERTS, R.S. An Assessment of Clinically Useful Measures of the Consequences of Treatment. *N. Engl. J. Med.*, 1988, 318(26):1728–1733.

LAURENCE, D.R., MCLEAN, A.E.M., et WEATHERALL, M. (éd.). *Safety Testing of New Drugs: Laboratory Predictions and Clinical Performance*. London Academic Press, London. 1984.

LINDGREN, B.W. *Statistical Theory* (2e éd.). Macmillan, New York. 1968.

LINSTONE, H.A., et Twoff, M. (éd.) *The Delphi Method*. Addison Wesley, Reading, Mass. 1975.

LOWRANCE, W. *Of Acceptable Risk*. William Kaufmann, Inc., Los Altos, Cal. 1976.

MAILHOT, R. The Canadian Drug Regulatory Process. *J. Clin. Pharmacol.*, 1986, 26: 232–239.

NAPKE, E. Canada. W.H.W. Inmann (éd.). Dans *Monitoring for Drug Safety*. MTP Press Ltd., Boston. 1986, pp. 49–69.

O'BRIEN, B. "What Are My Chances Doctor?" — *A Review of Clinical Risks*. Office of Health Economics, London. 1986.

OFFICE OF TECHNOLOGY ASSESSMENT. *The Impact Of Randomized Clinical Trials on Health Policy and Medical Practice: Background Paper*. U.S. Congress, Office of Technology Assessment, OTA-BP-H-22, Washington, D.C. 1983.

OFFICE OF TECHNOLOGY ASSESSMENT. *Assessing the Efficacy and Safety of Medical Technologies*. U.S. Government Printing Office, Washington, D.C. 1978.

OFFICE OF TECHNOLOGY ASSESSMENT. *Cancer Testing Technology and Sacharin*. Publication No. OTA-H-34., Washington, D.C. 1976.

PLISKIN, J.S., SHEPARD, D.S. et WEINSTEIN, M.C. Utility Functions for Life Years and Health Status. *Operations Research*, 1980, 28(1): 206–224.

ROYAL SOCIETY, THE. *Risk Assessment: Report of a Royal Society Working Party*. Royal Society, London. 1983.

SACKETT, D.L. et GENT, M. Controversy in Counting and Attributing Events in Clinical Trials. *N. Engl. J. Med.*, 1979, 131: 1410–1412.

SACKETT, D.L., HAYNES, R.B., GENT, M. et TAYLOR, M.W. Compliance. Dans Monitoring for Drug Safety W.H.W. Inman (éd.). MTP, Lancaster. 1980.

SACKETT, D.L. et TORRANCE, G.W. The Utility of Different Health States as Perceived by the General Public. *J. Chron. Dis.*, 1978. 31(11): 697–704.

SAMUELSON, W. et ZECKHAUSER, R. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty*, 1988, 1(1): 7–59.

SPETZLER, C.S. et VON HOLSTEIN, C.A.S.S. Probability Encoding in Decision Analysis. *Management Science*, 1975, 22:340–358.

SPILKER, B. *Guide to Clinical Studies and Developing Protocols*. Raven Press, New York. 1984.

TEELING-SMITH, G. (éd.). *Measuring the Social Benefits of Medicine*. Office of Health Economics, White Crescent Press Ltd., Luton, U.K., 1983, p. 5.

THOMPSON, M.S., READ, L.R., LEIGHTON J., et LIANG, M. Feasibility of Willingness-to-pay Measurement in Chronic Arthritis. *Med. Decision Making*, 1982, 4(2): 195–215.

THOMPSON, M.S., READ, L.R. et LIANG, M. Willingness-to-pay Concepts for Societal Decisions in Health. Kane, R. et Kane, R. (éd.). In *Values and Long-Term Care*. D.C. Health Publisher, Lexington, Mass., 1982, pp 103–.

TORRANCE, G.W. et FEENY, D. Utilities and Quality-adjusted Life-years. *J. Of Assessment in Health Care*. 1989 (sans presse).

TORRANCE, G.W. Utilities and Quality-adjusted Life-years. (en préparation). 1989.

TORRANCE, G.W. Measurement of Health State Utilities for Economic Appraisal: A Review. *J. Health Economics*, 1986, 5:1–30.

TORRANCE, G.W. Social Preferences for Health States: An Empirical Evaluation of Three Measurement Techniques. *Socio-Economic Planning Science*, 1976, 10(3): 129–136.

TREASURY BOARD OF CANADA. *Administrative Policy Manual*, chapitre 490, annexe E. Supply and Services Canada, Ottawa. 1979.

TUGWELL P., BENNETT, K.J., FEENY, D., GUYATT, G. et HAYNES, R.B. A Framework for the Evaluation of Technology: The Technology Assessment Iterative Loop. In: Feeny, D., Guyatt, G. and Tugwell, P. (éd.). *Health Care Technology: Effectiveness, Efficiency and Public Policy*. Institute for Research on Public Policy, Montréal. 1986, pp. 41–56.

TUGWELL P., BENNETT, K.J., SACKETT, D.L. et HAYNES, R.B., The Measurement Iterative Loop: A Framework for the Critical Appraisal of Need, Benefits and Costs of Health Interventions. *J. Chron. Dis.*, 1985, 38(4): 339–351.

VON NEUMANN, J. et MORGENSTERN, O. *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, N.J. 1944.

WALKER, S.R. et ASSCHER, A.W. (éd.). *Medicines and Risk/Benefit Decisions*. MTP Press, Lancaster. 1987.

WEBSTER'S NEW WORLD DICTIONARY (2e éd.). Simon and Schuster, New York. 1980.

WEINSTEIN, M.C. et FINEBERG, H.V. *Clinical Decision Analysis*. W.B. Saunders Company, Philadelphia. 1980.

WEINSTEIN, M.C. et STASON, W.B. Foundations of Cost-effectiveness Analysis for Health and Medical Practices. *N. Engl. J. Med.*, 1977, 296(13): 716–721.

WOLFSON, A.D., SINCLAIR, A.J., BOMBARDIER, C. et MCGEER, A. Preference Measurements for Functional Status in Stroke Patients: Inter-rater and Inter-technique Comparisons. In: Kane, R. et Kane, R. (éd.). *Values and Long-Term Care*. D.C. Health Publisher, Lexington, Mass. 1982.

ZBINDEN, G. Current Trends in Safety Testing and Toxicological Research. *Naturwissenschaften*, 1982, 69: 255–259.

Annexe A

Sélection de conférences récentes importantes sur la mesure de la qualité de vie par rapport à la santé

CONFÉRENCE	DATES	PARRAINÉ PAR	ENDROIT
Epidemiology: An Essential Tool for Drug Development	14–15 nov. 1988	Drug Information Assoc.	Charleston (Caroline du Sud)
Advances in Health Assessment Conference	13–15 juil. 1988	Institute of Medicine, Henry J. Kaiser Family Foundation	Menlo Park (Californie)
Workshop on Quality of Life and Cardiovascular Disease	7–10 juin 1988	National Heart, Lung and Blood Institute, Bowman Gray School of Medicine	Winston-Salem (Caroline du Nord)
Cost Benefit Assessment of Pharmaceutical Products	14–16 mars 1988	Drug Information Assoc.	Hilton Head Island (Caroline du Sud)
Measuring Quality of Life with Pharmaceuticals: Fundamentals and Techniques	10–11 déc. 1987	Pharmaceutical Manufacturers Assoc.	Baltimore (Maryland)
Issues of Quality of Life in Transplantation and Oncology	20 mai 1987	Netherlands Health Council, WHO Amsterdam Collaborating Centre on Quality of Life*, Netherlands Cancer Foundation	Institut du cancer des Pays-Bas, Amsterdam
Centre for Medicines Research Workshop	3 mars 1987	Centre for Medicines Research	CIBA Foundation, London
Measuring Quality of Life and Functional Status in Clinical Epidemiological Research	9–13 avril 1986	State of Science 1986	Sintra (Portugal)
Advances in Health Assessment Conference	19–21 fév. 1986	Henry J. Kaiser Family Foundation, the RAND Corporation	Palm Springs (Californie)

* L'*Amsterdam Collaborating Centre for Research and Training in Methods of Assessment of Quality of Life in Relation to Health Care* poursuit une série d'activités touchant le développement et la mise en application de programmes de recherche sur la qualité de vie des patients atteints de maladies chroniques.

Annexe B

Principales ressources bibliographiques

Centre d'information sur les index de santé

U.S. Department of Health and Human Services
Public Health Service
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

Ce centre d'information particulier a été établi en 1973 dans le but de promouvoir le développement des techniques du domaine maintenant connu sous le nom de qualité de vie.

Le centre d'information a deux fonctions principales. L'une est de publier une bibliographie sur les index de santé intitulée *Bibliography on Health Indexes*, qui peut être commandée au *U.S. Government Printing Office* pour la somme d'environ 2,50 $ par numéro. L'autre fonction est de servir de carrefour d'information entre les spécialistes qui mettent au point des évaluations de qualité de vie et les investigateurs cliniques; généralement entre le personnel du centre d'information et les investigateurs éventuels.

Annotated Bibliography for Risk Perception and Risk Communication

Centre for Research on Judgment and Policy
University of Colorado at Boulder
Attn : Doreen Victor
Campus Box 344
University of Colorado
Boulder, CO 80309-0344

Cette bibliographie annotée sur la perception du risque et la communication du risque est divisée en trois catégories principales : perception du risque, communication du risque et communication du risque en rapport avec la santé. Chaque section contient des annotations d'une grande variété d'articles de revues, de chapitres, de livres, d'études et de résumés de conférences se rapportant à sa rubrique. Cette bibliographie contient des références qui vont du début des années 1970 jusqu'au mois de février 1988 (25,00 $ US).

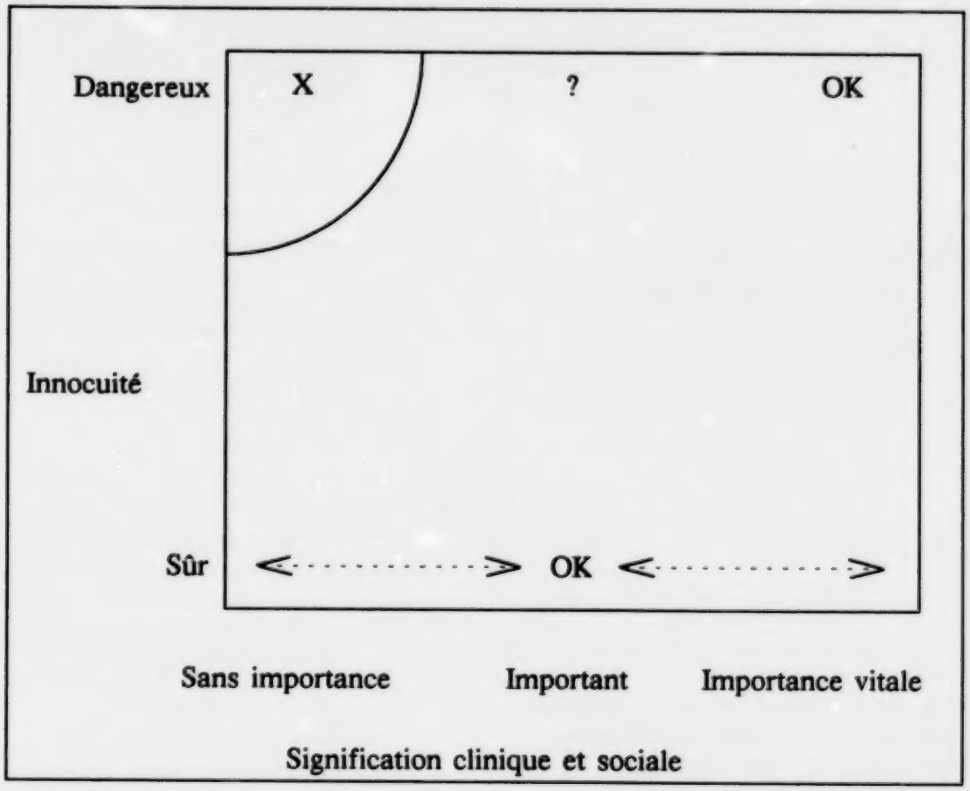

Figure 1 : **Matrice risques-avantages***

* Source : Teeling-Smith, 1983.

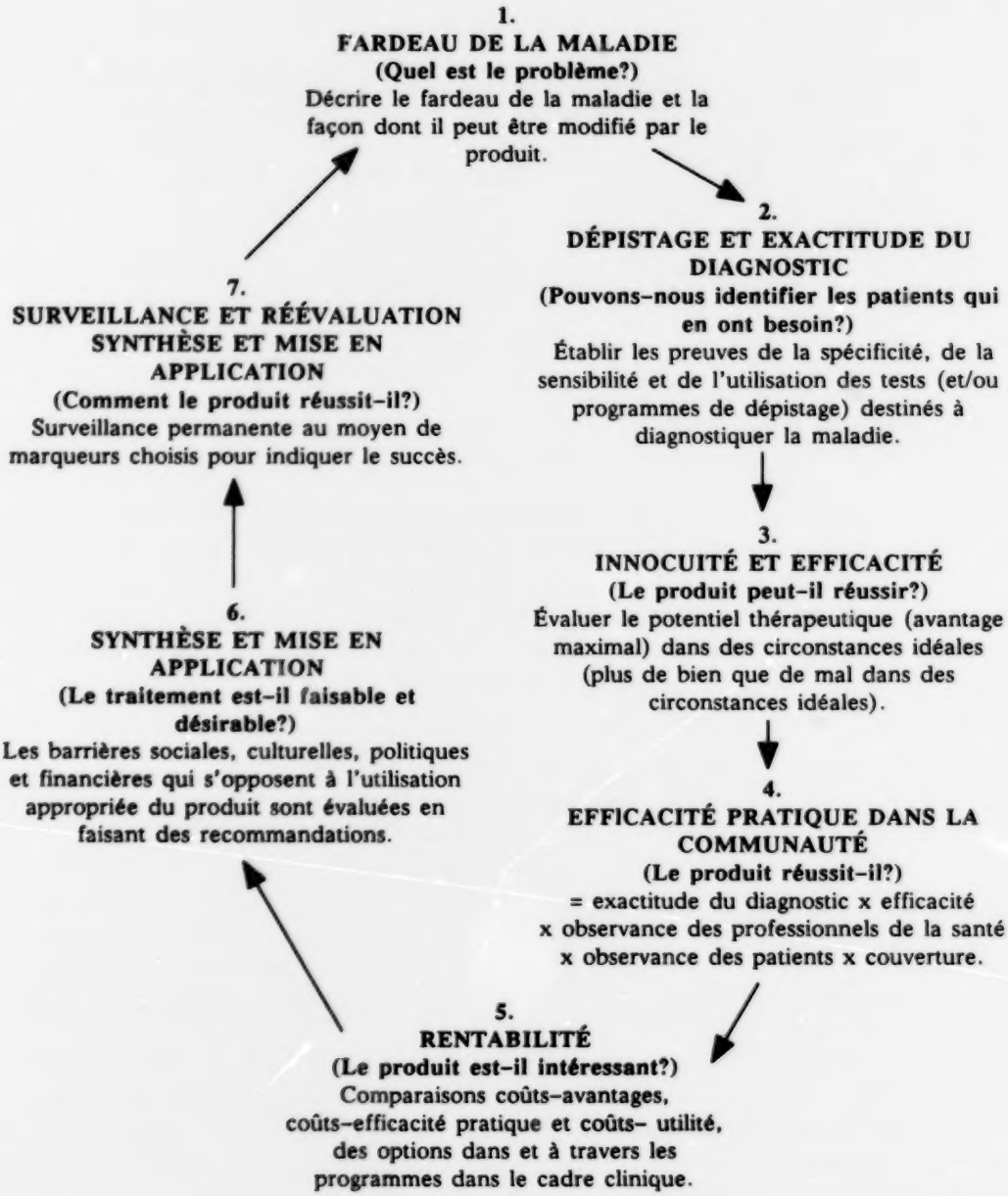

Figure 2 : **Boucle itérative de l'évaluation d'un médicament***

* Source : Adapté de Tugwell et al, (1985, 1986).

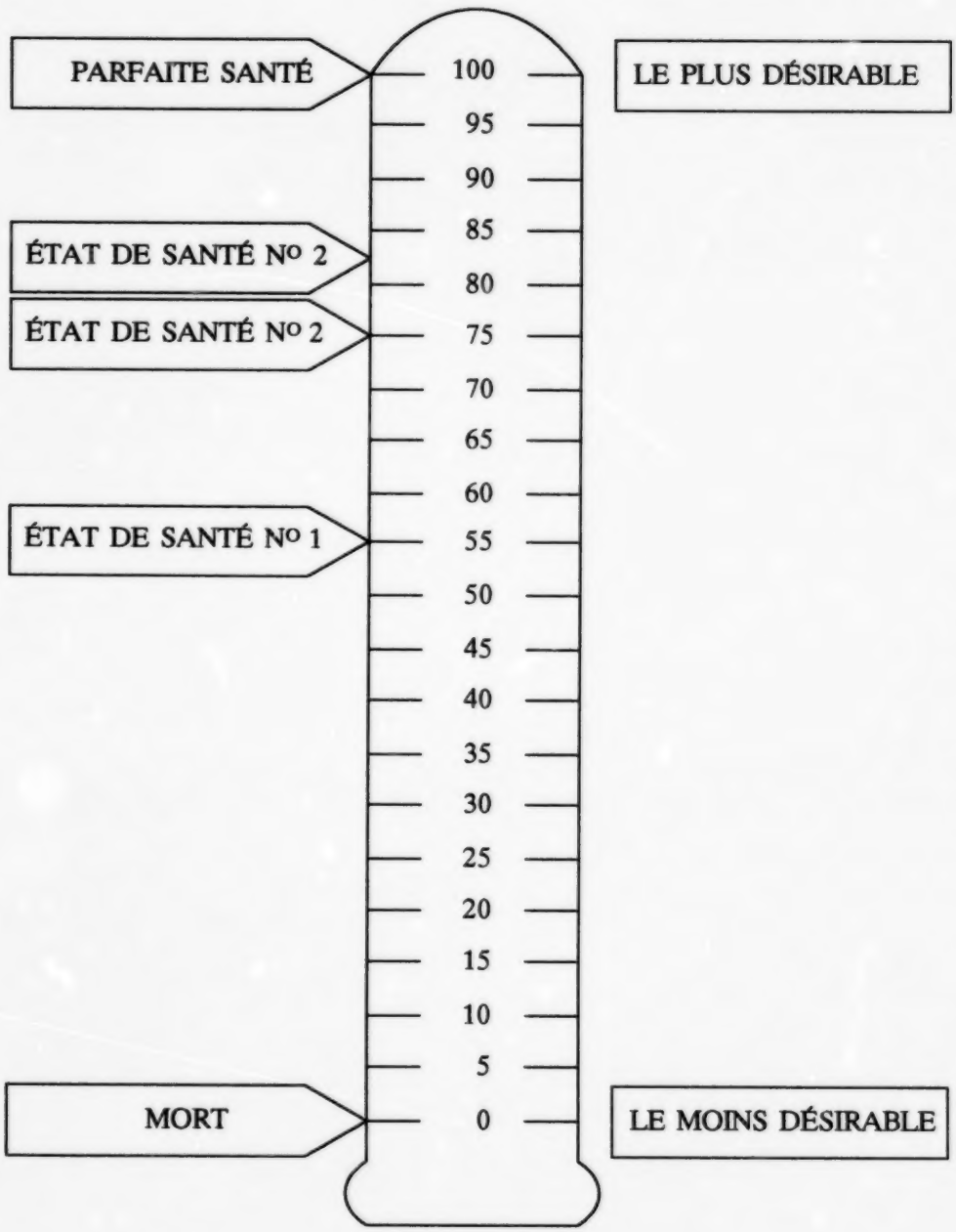

Figure 3 : **Utilisation du thermomètre des états de santé**

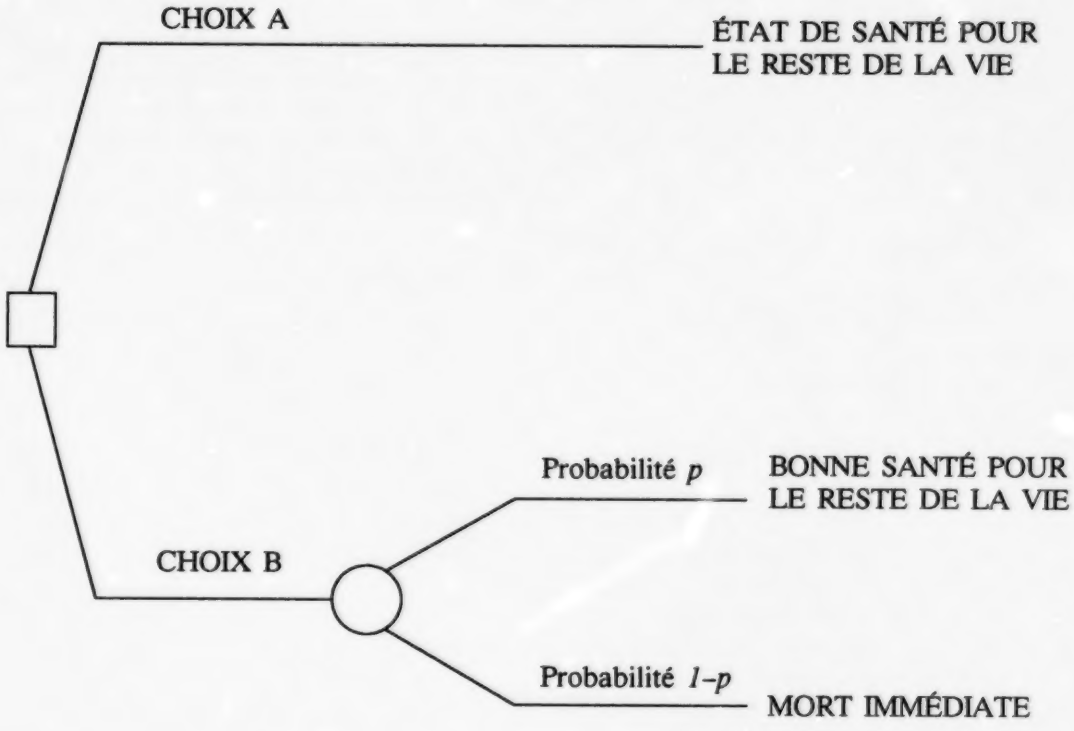

Figure 4 : **Méthode du pari type pour obtenir la valeur de l'utilité**

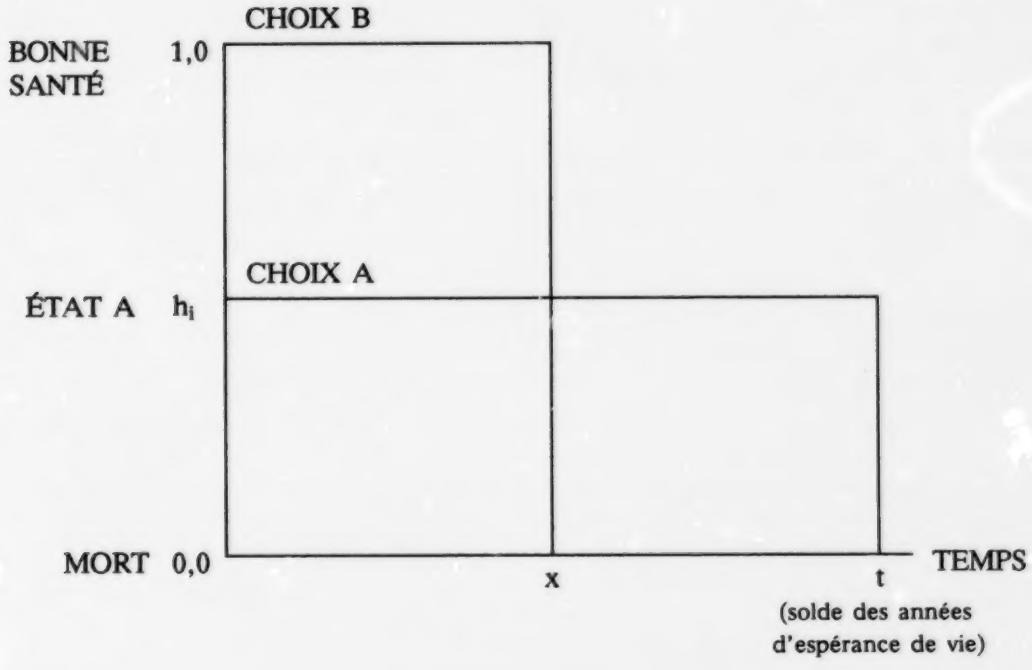

Figure 5 : **Méthode du choix de temps à vivre, pour obtenir la valeur d'utilité d'un état chronique**

Tableau 1 : **Quelques définitions de l' « efficacité »***

Source	Mot défini	Définition
Loi fédérale sur les aliments, drogues et cosmétiques	Efficacité, Efficacité pratique (interchangeables)	Un médicament est efficace s'il possède « l'action qui lui est attribuée, dans les conditions d'utilisation prescrites, recommandées ou suggérées dans l'étiquetage proposé ».
Cochrane dans *Effectiveness and Efficiency*	Efficacité, Efficacité pratique (interchangeables)	« L'effet d'une action médicale particulière se traduisant par une modification favorable de l'histoire naturelle d'une maladie ».
Organisation mondiale de la santé	Efficacité	Avantage ou utilité pour l'individu que présente un service, un traitement, un produit, une mesure de prévention ou de lutte, recommandé ou appliqué.
Discursive Dictionary of Health Care	Efficacité (variation de l'efficacité pratique)	« La mesure dans laquelle une action (ou des actions) diagnostique, préventive, thérapeutique ou autre (entreprise dans des circonstances idéales) atteint le résultat désiré ».
Office of Technology Assessment	Efficacité	La probabilité d'avantages que Technology des individus dans une population définie peuvent tirer d'une technologie médicale appliquée à un problème médical donné, dans des conditions idéales d'utilisation.
Service d'épidémiologie clinique et de biostatistique, Université McMaster	Efficacité	La mesure, le service ou le d'épidémiologie programme de santé fait-il plus de bien que de mal aux personnes qui se conforment exactement aux recommandations ou traitements associés?

* Adaptation de :

(1) Office of Technology Assessment, *Assessing the Efficacy and Safety of Medical Technologies*, Washington, D.C., U.S. Government Printing Office, 1978.

(2) Service d'épidémiologie clinique et de biostatistique, « How to read clinical journals: VII. To understand an economic evaluation (part A) », *Can. Med. Assoc. J.*, vol. 130, 1er juin 1983.

Tableau 2 : **Taxonomie des mesures de la qualité de vie***

Catégorie	Exemples
A. INSTRUMENTS SPÉCIFIQUES	Questionnaire sur les pneumopathies respiratoires chroniques Échelle des catégories d'arthrite Instrument de mesure de l'impact de l'arthrite
B. INSTRUMENTS GÉNÉRIQUES :	
B1. PROFILS DE SANTÉ	Profil d'impact des maladies Profil de santé Nottingham
B2. MESURES D'UTILITÉ	Pari type Durée de vie équivalente

* Source : Guyatt, van Zanten, Feeny, et Patrick (en préparation), *Measuring Quality of Life in Clinical Trials*, manuscript, 1988.

Tableau 3 : **Exemple de description d'un état de santé***

Capable de voir, d'entendre et de parler normalement pour son âge.

A besoin de l'aide d'une autre personne pour marcher ou se déplacer; a besoin également d'un équipement mécanique.

Occasionnellement agité, en colère, irritable, anxieux, déprimé ou souffrant de "terreurs nocturnes".

Apprend et retient les notions scolaires normales pour son âge.

Mange, se lave, s'habille et utilise les toilettes normalement pour son âge.

Exempt de douleur et d'inconfort.

* Source : Feeny, Barr, Furlong, Torrance, Rosenbaum, Weitzman et Horsman (en préparation), *Economic Evaluation of Treatment for Childhood Cancer: Costs, Benefits, and the Quality of Life*.

Tableau 4 : **Avantages et inconvénients des mesures de l'utilité***

A. Avantages

1. Possibilité de généralisation
2. Probabilité
3. Intégration des effets de morbidité et de mortalité
4. Capacité de capter deux dimensions de risque
5. Mesure des préférences pour des états de santé hypothétiques ou existants
6. Représentation de points de vue multiples
7. Incorporation des préférences en matière de temps à vivre et de risque
8. Bases théoriques rigoureuses
9. Intégration dans des mesures d'évaluation économiques
10. Intégration dans des essais cliniques

B. Inconvénients

1. Mesures d'utilité peu précises
2. Interviews demandant beaucoup de travail
3. Scores d'utilité peu faciles à comprendre
4. Mesures d'utilité qui ne peuvent être appliquées de façon universelle
5. Ne révèlent pas pourquoi une valeur d'utilité peut être élevée ou faible

* Source : D'après Feeny et Torrance, 1989.

Tableau 5 : **Variations des mesures cliniques de l'état de santé, et des indices composites entre le niveau de départ et le sixième mois, par groupe de traitement (N = 303)***

	Limites possibles (de pires à meilleures	Placebo (n = 149)	Auranofine (n = 154)	Effet du traitement (Valeur p)**
Mesures cliniques				
Nombre d'articulations sensibles		−4,5 ± 0,9***	−7,3 ± 1,0	0,01
Nombre d'articulations gonflées		−3,6 ± 0,7	−5,5 ± 0,7	0,01
Temps de marche 50 pieds(s)		+0,61 ± 0,66	−0,74 ± 0,40	0,11
Durée enraidissement matinal (min)		−12 ± 13	−28 ± 12	0,10
Force de préhension matinale (mm Hg)		−2 ± 8	+13 ± 8	0,02
Score composé		0,16 ± 0,05	0,35 ± 0,05	0,003
Mesures des fonctions				
Questionnaire d'évaluation de santé	3–0	−0,17 ± 0,04	−0,31 ± 0,05	0,01
Évaluation Keitel	98–0	+1,7 ± 1,0	−1,5 ± 0,9	0,003
Questionnaire qualité de bien-être	0–1	−0,001 ± 0,007	+0,023 ± 0,007	0,005
Score composé		0,05 ± 0,05	0,28 ± 0,05	0,001
Mesures de la douleur				
Questionnaire douleur McGill	78–0	−5,6 ± 1,0	−8,0 ± 1,1	0,02
Échelle de douleur (moyenne 6 jours)	0–10	+0,61 ± 0,13	+0,96 ± 0,18	0,09
Évaluation graphique de la douleur (ligne 10 cm)	10–0			
Score composé		−0,91 ± 0,18	−1,44 ± 0,17	0,01
		0,48 ± 0,07	0,74 ± 0,08	0,021
Mesures d'impression globale				
Arthrite :				
Échelle de catégorie	1–5	+0,31 ± 0,07	+0,65 ± 0,08	<0,001
Échelle (moyenne 6 jours)	0–10	+0,59 ± 0,14	+0,98 ± 0,18	0,11
Santé globale :				
Échelle, état présent	0–10	+0,39 ± 0,14	+0,65 ± 0,15	0,19
Échelle (moyenne 6 jours)	0–10	+0,35 ± 0,12	+0,87 ± 0,15	0,007
Évaluation de santé présente Rand	9–45	+0,51 ± 0,47	+1,82 ± 0,49	0,01
Ligne 10 cm, par le patient	0–10	+0,51 ± 0,13	+0,89 ± 0,16	0,1
Ligne 10 cm, par le médecin	0–10	+0,46 ± 0,12	+0,58 ± 0,13	0,2
Score composé		0,27 ± 0,05	0,50 ± 0,07	0,007

* Source : Bombardier, Ware, Russell, et al, 1986.

** Niveau de signification de l'effet du traitement déterminé par analyse de covariance avec ajustement pour les valeurs de départ : âge, sexe, clinique, catégorie fonctionnelle et variable testée.

*** Moyenne ± ETM

Tableau 6 : **Résultats des mesures de l'utilité et d'autres mesures***

	Fourchette possible (du pire au meilleur)	Placebo		Auranofine		Effet du traitement (valeur p)
		Valeur ou changement*	Nbre patients	Valeur ou changement**	Nbre de patients	
Mesures d'utilité						
Instrument de mesure d'utilité patient (score de variation dérivé)[1]	*0–100*	*9.9 ± 2.2*	*117*	*20.9 ± 2.6*	*122*	*0.002*+
Pari type (% de chance de décès)[2]	*100–0*	*30 ± 3*	*124*	*23 ± 2*	*119*	*0.07*+
Disposition à payer (% de revenu)[2]	*100–0*	*23 ± 1*	*105*	*21 ± 1*	*112*	*0.79*+
Autres mesures						
Questionnaire de dépression NIMH[3]	*60–0*	−4.1 ± 0.8	*149*	−3.3 ± 1.1	*154*	*0.54*++
Questionnaire d'état de santé général Rand[3]	*0–110*	−0.07 ± 0.75	*149*	+0.52 ± 0.79	*154*	*0.32*++
Mesures cliniques et fonctionnelles non comprises dans les scores composés						
Vitesse de sédimentation érythrocytaire (mm/h)[3]		+0.57 ± 1.68	*149*	−4.31 ± 1.6	*154*	*0.02*++
Questionnaire d'activités de la vie quotidienne de Toronto (score de variation perçu)[4]	*4–1*	0.00 ± 0.006	*149*	0.02 ± 0.007	*154*	*0.02*+

* Source : Bombardier, Ware, Russell et al, 1986.

* Moyenne + ETM.

\+ Niveau de signification de l'effet du traitement déterminé par analyse de covariance avec ajustement pour les valeurs de départ : âge, sexe, clinique et catégorie fonctionnelle, mais pas la variable testée.

++ Niveau de signification de l'effet du traitement déterminé par analyse de covariance avec ajustement pour les valeurs de départ : âge, sexe, clinique, catégorie fonctionnelle et variable testée.

1 Sans valeur obtenue au départ : dérivé rétrospectivement au cinquième mois des comparaisons faites par le patient entre le niveau de départ et le niveau actuel.

2 Sans valeur obtenue au point de départ : n'est pas un score de variation.

3 Valeurs obtenues au départ et au sixième mois.

4 Sans valeur obtenue au départ; observé rétrospectivement au sixième mois.